

LINKED DATA APPLICATION

Salma Sadek Jetlawei
salmajetlaei@gmail.com

Badriya Abdullah Altarhuni
badriyaabdullah@gmail.com

المخلص

في هذه الورقة نقدم المفهوم والمبادئ الفنية للبيانات المرتبطة، والتي تشير إلى أفضل الممارسات لنشر وربط البيانات المهيكلة على الويب، البيانات المرتبطة هي وسيلة لهيكل البيانات ونشرها على الويب وهي شكل من أشكال البيانات الموزعة على الويب التي يمكن الآلات من معالجة المعرفة ومشاركتها، مما يسمح أي شخص بنشر موارد البيانات وربطها ويمكن الوصول إليها ودمجها بكفاءة باستخدام HTTP وRDF، كما نناقش مفهوم متصفحات البيانات المرتبطة، والتي تسمح للمستخدمين بالتصفح في مصدر بيانات واحد والتنقل عبر الروابط إلى مصادر البيانات ذات الصلة، وتبسيط الضوء على تمكين أنواع جديدة من التطبيقات، مثل متصفحات البيانات المرتبطة ومحركات بحث البيانات المرتبطة، من خلال إنشاء مساحة بيانات عالمية تعرف باسم ويب البيانات.

ABSTRACT

In this paper defines the linked data as a set of best practices that are used to publish data on the web. Linked data is not just a field; it is occupying other fields as well and many technologies now most of the applications use linked data. Previously the data is in documented form but now every application needs linked data, mostly new applications need the data of the previous applications and they further implement functionalities to the previous data and make that data more effective this is done through linked data application, we also provide candid discussion relating to basic principles of the Linked Data, the technology stack of the Linked Data, the role of semantic query language as data processing pipeline, the role of the architectural aspects, and

some other approaches like testing and debugging approach and also about the Linked Data search engines .

Keywords: Linked data. Linked data application. Universal resource identifier. Resource Description Framework. Application Program Interface. Web of Data. Semantic web. Web Application. Linked Open Data.

1. INTRODUCTION

This paper is about the publishing structured data so the data can be used more effectively this technique is known as linked data. As the web is considered as global information space and now it is consisting of linked data instead of only linked documents, by this technique web data start getting mature through semantic web technology stack. As this is an essential need to have access to the raw data from the internet as previously there was no application but now linked data allowed to develop such applications [1].

In this paper, we have explored the history of the Linked Data, basic principles of the Linked Data, the technology stack of the Linked Data (HTTP, URIs, RDF, OWL), the role of semantic query language as data processing pipeline, the role of the architectural aspects, vocabulary -provides valuable sense to Linked Data, but also expresses the associations between data sources, the linking open data cloud, the function of the automated linking of the data set and some other approaches like testing and debugging approach and also about the Linked Data search engines. Furthermore, the linked data application following are the main modules which are mentioned below.

- Linked data browsers.
- Linked Data Search Engines and Indexes.
- Domain-specific Application.

- Testing approaches.
- Debugging approach.
- Linked data deployment.

1.1 Existing System

As in the need of integration of all data through the help of application which takes help of networking and the World Wide Web to make a linked data application. Furthermore, the basic needs developed big challenges for the applications should be integrated and for the applications and the technology used to maintain linked data. The linked data created many challenges in the development field as there are many applications which are developed to perform a certain task which a person cannot able to do but they have a flaw that their data in not linked. So, the basic flaw starts nowadays if an application does not have linked data as on societal level this can be done through semantic web technology to the web applications. The omnipresent and the use of applications raise a question to the privacy, and access to keep them ethically and socially updated. On the technical ground, the novelty of the linked data and the pace is increasing day by day. So, new applications developers are aware of the needs of the integration and this is managed and scaled properly to accommodate the users. Many algorithms and distributed system are placing their functionality and playing their role to overcome these new problems and the basic needs.

1.2 Drawbacks of Existing System

These are the few drawbacks of the existing systems.

Obsolete:

The existing system are not integrated and the data is neither centralized nor linked and are very tedious not exist at one place,

so these kinds of systems and the applications are getting obsolete nowadays.

Lack of Concurrency:

Previously on the web there was no such application was developed which have linked data but there was less application which has linked documents. As the previous applications, do not need integration or the linked data but as nowadays it is a basic need of any application so the existing system do not have concurrency and there is a need to develop those software's which have concurrency.

Not User-Friendly:

Most of the existing systems are not user-friendly. You need to access through the other PC to insert, update, delete or to access the data from any other place every time. It includes tedious, integration issue which explains that it is not user-friendly.

1.3 Problem Statement

In today's world, it is hard to find out where to travel when to travel and by which mean to travel. People have issues finding the option to travel per their time and budget in daily life. It is hard to find one platform to search schedules.

1.4.Solution

The solution, we need to use linked data browsers, linked data search engines, and their indexes and it is domain specific application which increases its need and processed to various testing phases and in this application, a new feature is added which is linked data deployment. Query processing is also needed in the mediator framework so that linked data can be able to integrate. There is a need of middleware framework to do scalable

management for the linked data. For the linked data service modelling, can be used for the internet of the things.

1.5. Features

This application includes following features.

- Linked data browser.
- Linked data search engines.
- Linked data indexes.
- Domain specific.
- Linked data deployment.
- Automated linking of data.
- All on one platform.

1.6 Advantage of Proposed System

- This application will help people to search data easily.
- The user can search by inputting information at one place and getting multiple outputs thus.
- A user can get information accordingly due to domain specification.
- This application use browser for searching.
- The application is user-friendly and provides ease to the users.

2.Related work

2.1. History of the Linked Data

The web has revolutionized the world into a new horizon of the linked data, through which the web is able, to create the links between the data from other sources; the system in linked data is heterogeneous, the data published through linked data system is machine readable, linked to an external set of data which establish the to create on of data sets [1].

Previously the data on the web was not wanted or was connected by other means, but now the barriers are overlapping through advancements in technology such as linked data [1]. The best data the information needs to be shared through linking there is a need to merge the data into a single place through the linked data application. It is revolutionizing the economic world and the world is producing more reliable work faster through the linked data.

Linked data is playing a vital role of the emerging science in recent years, the web totally changed from global information to the linked documents place [8]. The adoption or the evolution of linked data played its role in connecting the data through various places such as companies, people, television, radio programs [8]. The web of data provided new facilities for the developers to develop domain-specific applications previously web 2.0 mashups only worked with fixed data, but now there are a few linked data and search engines which crawl the data in the web and integrate it to provide more leisure same as in a local database. Linked data has enabled the generic and open nature in the field of web, as the web is growing wildly and enhancing linked data attributes along various strategies, linked data is leading to move itself towards a new horizon.

2.2The Linked Data and Technology Stack

Linked data is a vast technology which is used in almost every technology on the web. There are various names of the technologies which are also in use for the user interface of the applications [8].

There are two main technologies on which linked data relies one is hypertext transfer protocol (HTTP) and other is uniform resource identifiers (URI), URI provides identity to the generic

things which exist in the world basically it identifies the entities further than these looked up for the dereferencing, however, HTTP simple provide retrieving resources mechanism which was serialized for example photographs, furthermore the entities which cannot be shared by themselves on the network HTTP also retrieve their descriptions [8].

2.2.1 Resource Description Framework (RDF)

This technology is used to give services for the application which are used in base infrastructure and the databases to perform linked data successfully. Basically, RDF used to get graph-based model data to structure the linked data [10].

This is used to allow the client applications which navigate the data sources and the additional data as for being part of the linked data RDF links the entities with the other data sources because data sources provide the information so sometimes semi-automated or automated techniques are used to generate the links for RDF [10].

2.2.2 Web Ontology Language (OWL)

Classes and the data are managed through web ontology language and the properties of the linked data are handled. It is used to develop ontologies [8]. It basically provides the base to create vocabularies so that it can be described as the entities and the relation between them, then the vocabulary is expressed in the resource development framework.

They are used to explain the classifications of the networks and the taxonomies which were defined as the structure of different domains such as classes, and the verbs [12].

2.2.3 Semantic Query Language (SPARQL)

In the linked data, it is used as data processing pipelines. In the technology stack, it will be used as query generator and uses sparql-client and eea. sparql.

This is a service which provides a legacy of the HTML documentation for the web which does not contain the RDF data so it enables the extraction of the linked data. The service provides by giving queries through which further entities changed into RDF graph [10].

This is semantic query language which manipulates and retrieves the stored data from the resources description framework and this is the main feature of the semantic web it further allows the query into three parts first is the conjunction, disjunction, and the optional pattern.

2.2.4 Hyper Text Transfer Protocol Secure (HTTPS)

Hypertext Transfer Protocol Secure is used to communicate in the linked data. HTTP is an application layer protocol which was previously used but now HTTPS is used as the major change in this is that HTTP is not secure however HTTPS is secure so it's reliable to be used in the application.

RDF is used as data model and however, HTTPS is used as data access mechanism which further compares with the API's and fetches interfaces and the heterogeneous data model [12]. The web of data is an additional layer which overcomes the documented web previously.

2.2.5 Universal Resource Identifier (URI)

Universal Resource Identifier is used in the applications as identifiers and then these identifiers are referred to the XML [4].

The basic concept of the linked data and technology stack is shown in figure 1.



Figure 1: The Linked Data Technology Stack Diagram

2.3 The Linking Open Data Project

In early stages, the participation was not that much but now it is an emerging field of computer science. As this project is growing and the growth is increasing and openness of the project can be judged that anyone can participate in this by just publishing the data set along with interlinking the datasets and the published data sets should be per the linked data principles [13]. Distinct data set are further explained and the range of the web of data and scale which originate the openness of the linked data is further explained. Now that, big organizations are significantly involved in linked data, in this diagram, each node represents the distinct data set which is published as linked data.

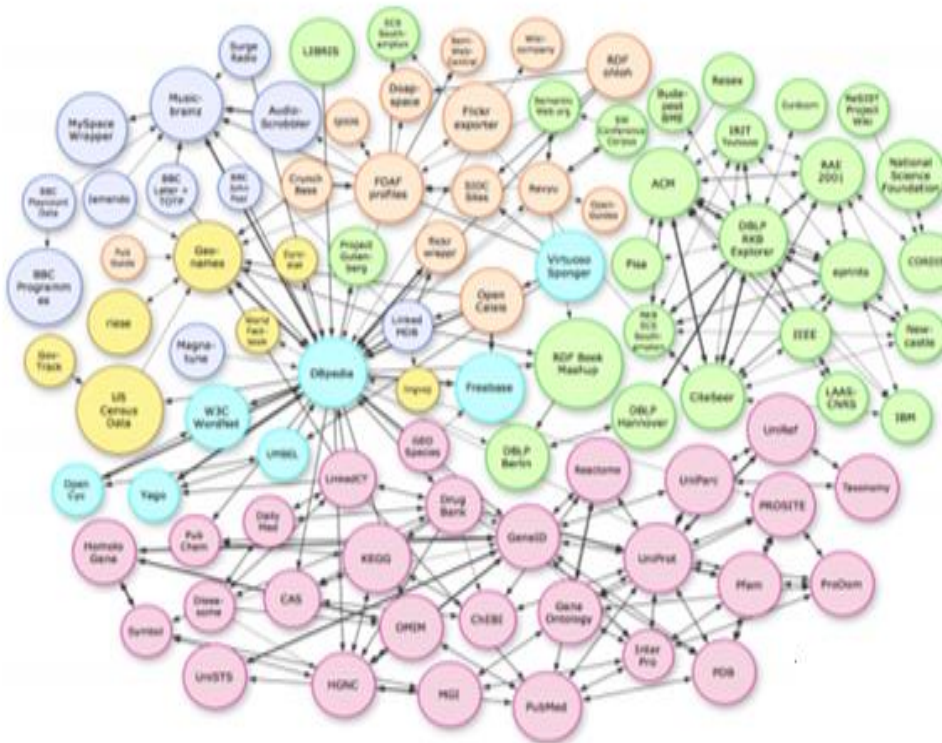


Figure 2: The Linking Open Data Cloud Diagram adapted from (Bizer et al., 2009).

In figure 2, it is shown that the data set servers are linking as hubs in the web, RDF and URI is provided as two data sets for the most common concepts or the entities, furthermore, they are referenced to the most specialized data sets which have developed as hubs through numbers of the data sets are connected [13]. The architecture of linked data is considered in startup few web crawlers are built to interlink. In the figure 2, it is clearly shown that the linking open data contains cross domains, government, geography, life science, media, linguistics publications, social

networking and user generated [6]. The things which are highly considered for the application is not just being created through the help of the built-in libraries, the application should use the benefits of the RDF. Secondly, the application should also follow the baseline criteria and real-time data should be used [6]. Thirdly the scientific paper should also be submitted which explains the major aspects and the procedure of implementation, which helps to understand the purpose of the application. Every day the challenges are changing multiple times and trends are also changing, so compare it with them as well [10].

3. PUBLISHING LINKED DATA ON THE WEB

Implementing linked data principle by publishing the linked data on the web, the data providers insert the data to the global data spaces through which data get discovered and are used by the applications.

3.1 Basic Principles of Linked Data

In this, links have the relationship anchors in the hypertext language and the documents are written in the HTML, then the linked data is described by RDF and URI explains the concept and the object [15].

The basic process has three main steps.

1. Set the RDF Links:

On the web, set the links to the other data sources, through which the users can reach to the linked data.

2. Assign the URI:

In the RDF presentation, dereference the URI on HTTP and assign the URI for the entities.

3. Provide the metadata:

Provide the metadata of the published data through which the users can access the published data and can choose different means for access.

3.2 Relevant Aspects of Web Architecture

The limitation sets in the architectural style which resists the features and explains the role of the architectural aspects [15].

RDF-based application realized the aspect that there are numerous types of the architectures which are available. Besides this, the other surveyed applications which impose limitations on the web services and their infrastructure can be used to build the application so the application depends on the limitations implemented with respect to the architecture chosen. Most of the applications use the service-oriented architecture for the linked data applications; the applications which use layered approach mostly reuse resources such as the databases or any middleware functionalities [2].

In the component based architecture, most of the application depends on the reusability of the tools and the libraries as the components. The architecture with very less denominator, in this application we will use component based architecture for our concepts it let us allow to perform high-level functionalities which are required for the linked data for the web of data and for the semantic web[7].

3.3 Drawing on a Practical Linked Data Scenario

Different processes are under laid which describes about the linked data and the flowchart of the application is then described. Figure 1 shows the overview and the scenario of the linked data with its overall working and generation process, where as in this

figure the conductive tasks and the non-consecutive tasks and the outputs are also shown.

The linked data process is basically divided in to eight further sub parts. At first, the data source is introduced to the linked data and then the data source is accessed, then the license is analysed, furthermore the source of domain name is specified and then data source is analysed, then ontology is created and data is sent as RDF format and thus the generated data is linked with the other linked data [2].

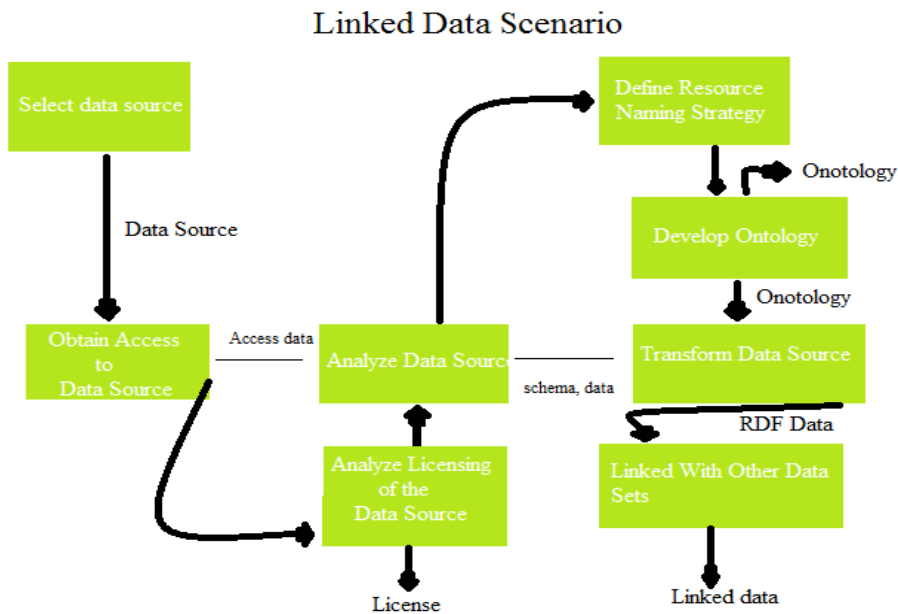


Figure 3: Linked Data Scenario

3.4 Testing and Debugging Approaches for the Linked Deployment

In this section testing is done on the complete integrated system it is done to check the systems compliance with the

requirements of the system the scope of the testing is checked through black box testing which check the algorithms and the code and the inner design and structure of the code [14].

Testing phase is done through quality assurance these are the set of different activities performed on the application to evaluate the product which is developed. Further the standards of the software are checked through software quality control function which checks linked data application complete processes and the procedures step by step and verify the products internally and externally [5].

After developing linked data application there are few test on the application needs to be done which are done in the black box and are known as specification testing. The further testing activities are performed in this phase and through this, program errors, and program requirements are observed.

The main purpose of the testing is to check the working of the linked data application.

1. Specific component behavior error is checked.
2. Specific performance error of the application is checked.
3. Specific function errors are thus checked to improve the working of the application.
4. User based error are also checked.
5. Observing all above then black box error is checked in linked data application.

4. LINKED DATA APPLICATION

Various applications are being built nowadays and research effort is taking place on the linked data, a huge number of the linked data is also publishing on the web. Explaining more linked data application further divided into three main categories, linked

data browser, linked data search engines and indexes, and domain specified.

Mainly the things to be focused are the characterization of the application, architecture of the linked data application; the development framework, and web API's. According to the rules what completes the linked data application are,

- Consumes Linked data.
- Manipulates and produces linked data.
- Web application.

Summing these three things/functionalities gives us a linked data application as explained in figure 1.

In figure 1, it is shown that mashups are considered which consume linked data it does not mean that sources contain RDF data. The app mostly uses scrappers to convert data into the linked data. Manipulates and produces linked data basically it only updates RDF data and gives accessibility to the web of data [3]. Web application gives the same features as in desktop and the mobile application performs integration to export data sometimes operated on the web as well.

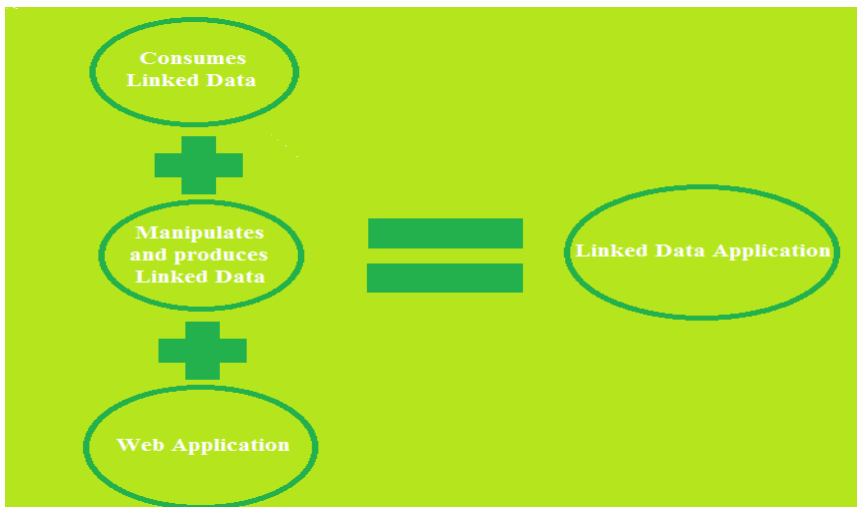


Figure 4: Three Main parts of Linked Data Application

There are various types and levels of the linked data application extrinsic it uses the semantics, intrinsic is uses conventional technologies, consuming it is wrapper based or it retrieve from the source [3], producing it publishes the linked data, shallow it uses RDF, strong it uses OWL and high level of the representation, isolated it create its own vocabularies, integrated it involves reuse of the information.

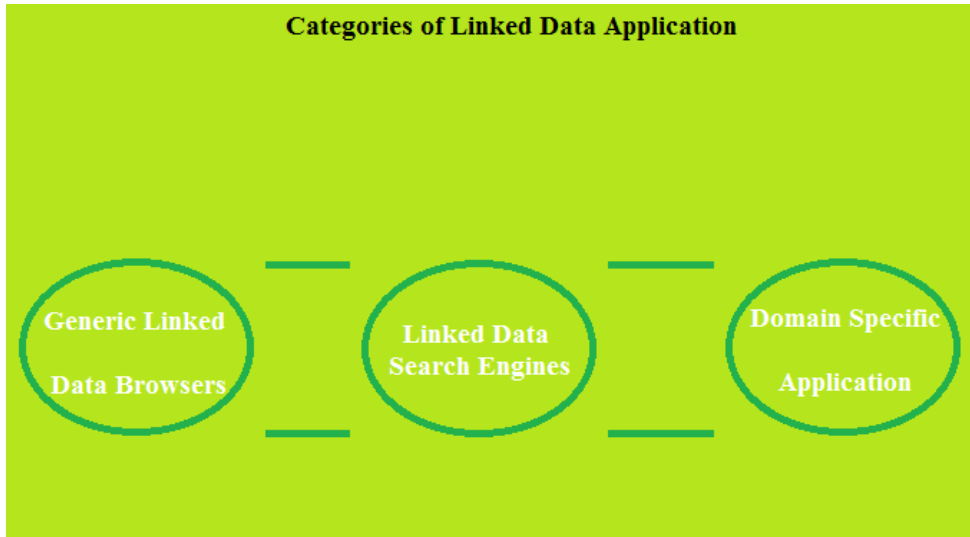


Figure 5: Categories of Linked Data Application

4.1 Linked Data Browsers

The normal web browsers only allow the user to search HTML pages through the hypertext links but in the linked data browsers user is allowed and can search the data sources by following the links, which are known as RDF triples.

For example, in RDF DBpedia a user searches for a city which follows its birthplace which further also refers to the city of the birthplace, this approach is also used in the BBC broadcasting [4]. It leads to the result that the user can start navigating the data

source and thus traverse the web by RDF and the city name, thus it follows the approach of hyper data browser which can be visualized as the web of data for the hypertext navigation along as the direct application.

The data further provides an interface which then provides the opportunity and the challenges except for hypertext mark-up language. Analyze the data link by link or either takes a bulk to browse the web of data for exploring. For traversing the tabulator (berners-lee et al,2006; berners-lee et al,2008), it basically allows the user to browse the data for traversing and the outline is also shown to query about the interest or the query about any other pattern represented in the web of data [10]. After querying the results are then shown and analyzed with the conventional data and the presentation data or the methods used for presenting data i.e. maps, browsers, walls, and timelines etc.

Tabulator and Marbles (becker and bizer, 2008) are the browsers to track the data, during the emergence of the data from sources which is shown in figure6 [4].

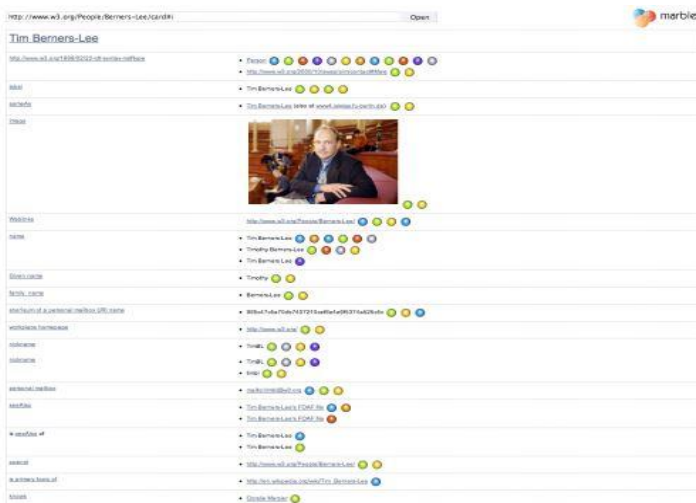


Figure 6: showing emerged data from different sources

4.2 Search Engines Linked Data and Indexes

In the documented web of the hypertext mark-up language, web browsing and the searching are the two prominent factors for the interaction for the web of data. The browsers help to find the information and provide a technique to process navigation nowadays many linked data search engines are working by following the RDF links to provide the user an aggregate web of data. Furthermore, it is divided into two major parts search engines which are human oriented and indexes which are application oriented.

5.CONCLUSION

The paper is to publish linked data in this paper history of the linked data, the linked data technology, the linking open data, principle of linked data, aspects of the web architecture furthermore consumption of linked data, practical linked data scenario, architectural approaches to publishing linked data choosing URLs and vocabularies to identify and describe resources, deciding what data to return in the description of the resource on the web and methods and frameworks for automated linking of data sets are explained [12].

The web is being used everywhere and basic need we feel affection towards the way it's developing rather than a simple web of linked documents in past. As typify data on the web is clear and obvious, but its next step is slightly harder. Everyone comes to the web in their own way but when it comes here for Linked Data, makes them closer. The sets of techniques play a vital role in representing and combining prearranged data on the web. We have already discussed that how can we connect, access, share and utilize Linked Data as well as we have seen its remarkable property that

it can be combined not only with other data but also with the Linked Data to shape new knowledge.

To be sure that the set of techniques of Linked Data make us able to share information, and our knowledge with others and on the web it may be shared, discovered and connected with other people's data. It is a property of Linked Data that it freezes data or information from a proprietary container so everyone can easily use it. Literally, anything can be represented by this technique.

Nowadays, millions of websites are using this practice (Linked Data) to depict their pages. As, this method, in addition, is also utilized to add metadata to web pages and that format is called schema.org, and it is supported by many prominent companies e.g. Microsoft, Google, Yandex, and Yahoo. This format permits all sorts of data (e.g. images, knowledge, reviews and events) to be added to a web page. So, today, this format is creditable in this context also.

What will be the future of Linked Data?? Per my view, Agile standards and incremental efforts as schema.org will be popular and as aligning and cleansing of data is a common issue in the case of big data, In the future, machine learning way will be able to mechanize such method to build smart Linked Data clients which have competent of aligning and to cleansing the raw Linked Data. And we will see the unification between several heterogeneous platforms like wikidata, the Linking of Data clouds, HTML pages. In near future, we will find it user-friendly as well.

As, we have come to know that Linked Data is a recommended best practice to connect the related data that was not previously linked by using the web and it makes lower the hindrance for

sharing and attaching pieces of data, knowledge, and information. So the value of Linked Data in information technology is worthy. Here I quote some words of a critic;

“While big data will make an organization smarter and more productive, open and linked data will make it richer”

It’s undeniable that Linked Data plays a vital role to solve some important problems that always made frustration for the IT industry for many years in past. This technique allows its user or organization to publish more, reusing more and combine more data.

According to Michael A. Keller

“Linked Data could provide the antidote to the chaos and complexity of the current overabundant array of Too simple search mechanisms with too little precision and too short recall of relevant results”

6. REFERENCES

1. Amit P. Sheth, (2011). Semantic Services, Interoperability and Web Applications: Emerging Concepts, 214-298.
2. Necula, S. C. (2020). Semantic Web Applications: Current Trends in Datasets, Tools and Technologies' Development for Linked Open Data. Informatica Economica, 24(4).
3. Heath, T., & Bizer, C. (2022). Linked data: Evolving the web into a global data space. Springer Nature.
4. Benjamin, Cyganiak, Conor, Decker, Heitmann, Richard Hayes, Stefan, (2014). Architecture of the linked data applications.
5. Heath, T., & Bizer, C. (2022). Linked data: Evolving the web into a global data space. Springer Nature.
6. Ristoski, P., Rosati, J., Di Noia, T., De Leone, R., & Paulheim, H. (2019). RDF2Vec: RDF graph embeddings and their applications. Semantic Web, 10(4), 721-752.

7. Zeng, M. L., & Clunis, J. (2020). FAIR+ FIT: Guiding Principles and Functional Metrics for Linked Open Data (LOD) KOS Products. *J. Data Inf. Sci.*, 5(1), 93-118.
8. Bonatti, P. A., Decker, S., Polleres, A., & Presutti, V. (2019). Knowledge graphs: New directions for knowledge representation on the semantic web (dagstuhl seminar 18371). In *Dagstuhl reports* (Vol. 8, No. 9). Schloss Dagstuhl-Leibniz-Zentrum fuer Informatik.
9. Rasmussen, M. H., Lefrançois, M., Schneider, G. F., & Pauwels, P. (2021). BOT: The building topology ontology of the W3C linked building data group. *Semantic Web*, 12(1), 143-161.
10. Houssein, E. H., Ibrahim, N., Zaki, A. M., & Sayed, A. (2022). Semantic Protocol and Resource Description Framework Query Language: A Comprehensive Review. *Mathematics*, 10(17), 3203.
11. K. Bollacker et al. "Freebase: a collaboratively created graph database for structuring human knowledge". In: 2008 ACM SIGMOD international conference on Management of data. ACM. 2008.
12. Elstermann, M., Wolski, A., Fleischmann, A., Stary, C., & Borgert, S. (2021). The combined use of the web ontology language (OWL) and abstract state machines (ASM) for the definition of a specification language for business processes. *Logic, Computation and Rigorous Methods: Essays Dedicated to Egon Börger on the Occasion of His 75th Birthday*, 283-300.
13. Nikolov, A., et al. (2008). Integration of Semantically Annotated Data by the KnoFuss Architecture. *Proceedings of the 16th International Conference on Knowledge Engineering and Knowledge Management*.
14. Radu VasIU, Silviu Vert, (2014). Relevant aspects for the integration of the linked data in mobile augmented reality applications for tourism. 334-345.

15. Hyvonen, E. (2022). Publishing and using cultural heritage linked data on the semantic web. Springer Nature.
16. Allemang, D., Hendler, J. A., & Gandon, F. (2020). Semantic web for the working ontologist. ACM Press.
17. Polleres, A., Kamdar, M. R., Fernández, J. D., Tudorache, T., & Musen, M. A. (2020). A more decentralized vision for linked data. *Semantic Web*, 11(1), 101-113.
18. Hyvönen, E., Leskinen, P., Tamper, M., Rantala, H., Ikkala, E., Tuominen, J., & Keravuori, K. (2019, May). BiographySampo—publishing and enriching biographies on the semantic web for digital humanities research. In *European Semantic Web Conference* (pp. 574-589). Cham: Springer International Publishing.