International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878, Volume-8 Issue-3, September 2019

De-Centralized Hash URI Procedure for Making Digital Artifacts as Immutable, Reliable, Verifiable on Web

Naseema Shaik, E. Sreenivasa Reddy

Abstract*— Main vision behind semantic web data maintenance is to make the text or content of web is interpretable, among different things, and sophisticated search procedures over huge amount linked web data. In this case different human beings perform fraudulent text which can be found on internet web. So that main research concentrated on automated approaches autonomously identified and analyzed content present in web. Making digital artifacts on web in terms of text, code and others verifiable and reliable, because of general nature of digital artifacts i.e immutability, it is a serious concept to produce reproduces the updated results of processes that display modified content in web resources. So that in this paper, we propose De-centralized Hash URI approach (DHURI) to solve the immutable concept of digital artifact generation in nano-publication with unique hash based cryptography value. In this approach, we provide Decentralized secure URI for identification of digital artifacts with parallel structure representation of data such as nano-publication. Basic concept of this approach is to handle digital artifact immutable concept with evaluation of existing approaches, and also describe the performance of proposed approach in de-centralized framework.

Index terms: Uniform resource identifier, digital artifacts, verifiable, immutable, de-centralized architecture, nano-publication and resource description framework (RDF).

I. INTRODUCTION

Main vision behind semantic web with substance of machine capable with different notations, permitting, in addition to other things, for computerized collection and modern inquiry systems over a lot of connected information. As even human clients are now and then simple to trap by spam and deceitful substance that can be found on the web, we ought to be considerably progressively worried on account of mechanized calculations that self-rousingly investigate semantic web content. Without suitable counter-measures, pernicious on-screen characters can damage or control such calculations by including only a couple of painstakingly controlled things to huge arrangements of input information. To address a portion of these issues, various logical information vaults have showed up, for example, Figshare and Dryad (http://figshare.com, http://datadryad.org). Besides, Digital Object Identifiers (DOI) has been supported to be utilized for articles as well as for logical information (Paskin, 2005). While these methodologies unquestionably improve the circumstance of logical information, specifically when joined with Semantic Web procedures, they have by the by various disadvantages: they have brought together models, they give us no probability to check whether the information have been (intentionally or inadvertently) altered, and they don't bolster get to more regularities from data sets, (for example, singular information passages). We contend that the incorporated idea of existing information vaults is conflicting with de-centralized ordinary with different scientific relations, and that it has genuine results concerning unwavering quality and trust. The associations running these stages may eventually fail, be gained by financial specialists who don't feel focused on the standards of science, or for different reasons become incapable to keep their sites ready for action. Despite the fact that presented information achieves guarantee relations on data sets with accessible approaches to test that hash value is dependable or not.

In particular, our exploration question is: Can we make a decentralized, solid, and reliable framework for distributing, recovering, and filing Linked Data in the type of sets of nano publications dependent on existing web norms and framework? It is critical to note here that the word reliable has a wide importance and there are various types of trust included with regards to recovering and utilizing datasets from a few outsiders. When investigating existing datasets, a particular sort of trust is expected to choose regardless of whether an experienced dataset is fitting for the given reason. An alternate sort of trust is expected to choose whether an acquired record accurately speaks to a particular adaptation of a particular dataset that has been utilized. Just the second sort of trust can be accomplished with a specialized arrangement alone, and we utilize the word dependable in this paper in this limited specialized sense covering the second sort of trust.

So that in this paper we propose and present De-centralized Hash URI approach (DHURI) to solve the immutable concept of digital artifact generation in nano-publication with unique hash based cryptography value. In this approach, we provide Decentralized secure URI for identification of digital artifacts with parallel structure representation of data such as nano-publication.

Manuscript published on 30 September 2019
* Correspondence Author
Naseema Shaik*, Research Scholar, Acharya Nagarjuna University, Guntur, AP, India, Email: meetmsnaseema@gmail.com
Dr. E. Sreenivasa Reddy, Dean, Faculty of Engineering, Acharya Nagarjuna University, Guntur, AP, India

Proposed approach defines in this paper describes trusty URI digital artifacts which are relates to immutable and describes advance changes in URI. However this approach also represents RA and FA in generated URI with different notations and it's hash value representation. Experimental results give efficient performance of proposed approach with respect to time and hash code generation of URI on semantic web.

II. DE-CENTRALIZED HASH URI APPROACH (DHURI) IMPLEMENTATION PROCEDURE

This section present and describe the procedure of De-centralized Hash URI approach (DHURI) builds upon the kinked data on web to linked RDF representations. This approach deals with structure data assume that that data already present in RDF format. We exploit the different data sets present in RDF divide into small pieces w/t any changes in semantic web data. Specifically, our proposed approach addresses the nano publications and describes the trusty URIs procedure; it is the basic representation to handle reliable implementation on semantic web applications based on follow-up noise principles.

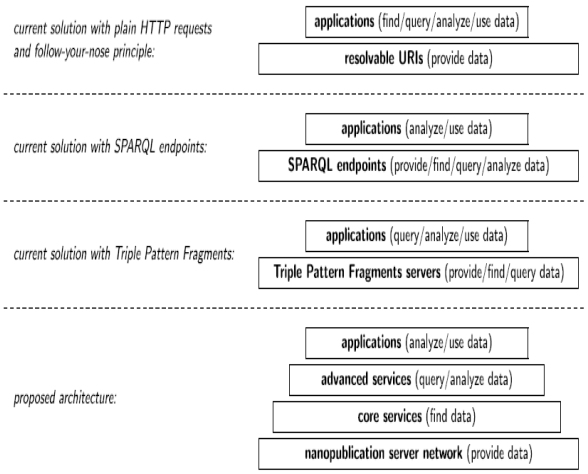

Fig 1 Semantic web nano publication representation of proposed approach.

Main description behind implementation of semantic web application appeared and described in figure 1, main option i.e. HTTP GET request references present in URI and eliminate noise i.e. unrelated data in resolvable URI dependent on application which are relates to find important assets, enquires in hash values present in outcome URI. This methodology describes the semantic web oriented applications with tree-traversal which describes effectiveness and fulfillment of URI generation.

a. Server Communication to Publish Nano publications

Basic implementation of implemented approach is described above with low-level data communication, here we present de-centralized nano-publication server relation with REST API to describe the procedure of nano-publications. Main presentation behind nano-publication is dependability of system implementation required trusty URI identifier i.e. these URIs must be changed on user side before stored into distributed server communication. It presents three components publication, retriever and distribution, nano-publication server associate with another server which are consists duplicate nano-publications and also enable file transfer operations on clients with new nano-publications which are conveyed with distributed framework, semantic relation between server and user with median server communication described in figure 2.

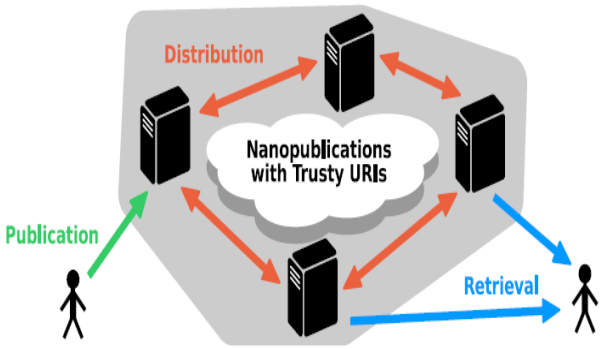

Fig 2 De-centralized representations for Nano publications via servers.

Basic substance of proposed framework with generation of trusty URI which consists architectural implementation with different nano-publications. Main advantage behind that generation of nano-publication makes simple and evaluate the process in generation of hash value and RDF data base. Sub-sequent advantage of servers is to manage identifier on board with trusty URI which consists exceptional description in implementation, trusty URLs are change or not with potential communication in hash URI. This implies servers just need to manage including new sections yet not with refreshing them, which takes out the difficult issues of simultaneousness control and information trustworthiness in disseminated frameworks. (Similarly as with old style distributions, a nano-publication once distributed to the network cannot be erased or "unpublished," yet just checked withdrew or on the other hand supplanted by the production of another nano-publication.) Together, these viewpoints essentially streamline the structure of such a system and its synchronization convention, and make it solid and effective even with constrained assets.

b. Index based Nano-publication

One of the center thoughts behind nanopublications is that every one of them is a little nuclear bit of information. This suggests investigations will for the most part include something beyond one nano-publication and commonly countless them. So also, most procedures will create something beyond one nano publication, potentially thousands or even a great many them. In this manner, we should almost certainly assemble nanopublications and to recognize and utilize huge accumulations of them. Given the flexibility of the nano publication standard, it appears to be clear to speak to such accumulations as nanopublications themselves. Accumulation of data presentation in nano-publication with different notations contain previous enormous for huge accumulation and describe their properties relates to nano-publications.

Nano publications can't contain unrelated trusty URIs which enable reference interfaces to accentuate not accumulations in outsourced user data.

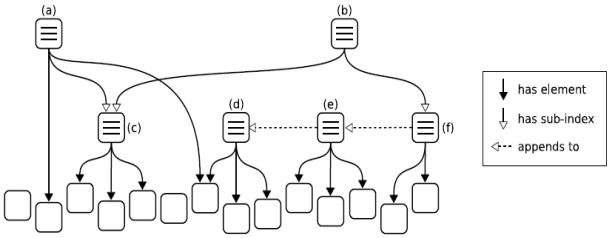

Fig 3 Formation of nano-publication using hash index with respect to acquiring data publishing.

Reference label contains total nano-publications to become very enormous, to describe all the records present in nano-publications in size and other quantity references and also helps in self assertive in size with different permit lists by different lists which consists nano-publications. We present 1000 nano-publications with different possible labels, however, all the users' accounts present in trusty hash URI classified with different hash URIs for different files each other. For example, in 100000 nano-publications, classify 100 files from nano-publications with expected data presentation for all the users' accounts with sorted relations based on reference classes. Each record sorts relations based on classified references with sub-part present in trusty URI with different notations and finds the sub file representations.

Above relations are classified based on structure of nano-publication which is described in figure 3, it describes total representation of trusty URI which contains sub part of file i.e. FA and RA, main extreme operations with sub-records which represents different class labels. If any nano-publications does n't contain references classification label then it describes the degree of publications based on general procedure to identify existing nano-publications which are relates to distributed with recovered nano-publications.

c. Trusty URI Publishing

Based on above procedure discussed in above sections, let us consider the expletory approach to handle and describe the general concept to store and create index based URI for http link on semantic web data. To describe the procedure to create trust URI using some individual steps on developed user interface. General structure of the index based trust URI with Base 64 hash value notation is

http://example.org/r1.RA5AbXdpz5DcaYXCh9l3eI9ruBosiL5XDU3rxBbBaUO70

Shown in above URI after r1 shows the base hash value with 64 bit link, in this first two characters like RA describes base values with consist information about the data and its version, remaining characters describes its hash value. Basic representation of precise specification of generated URI could be found in semantic web data.

Our proposed approach is very compatible and support with ni-URI's, and transform the relations between hash based URI with ni-URI or without ni-URI explicitly specify the authority

ni:///sha-256;5AbXdpz5DcaYXCh9l3eI9ruBosiL5XDU3rxBbBaUO70

ni://example.org/sha-256;5AbXdpz5DcaYXCh9l3eI9ruBosiL5XDU3rxBbBaUO70

If the module identifier changes doesn't effect on uniqueness of the generated hash value, to develop this, in this approach, we use augmented module description i.e.

ni:///sha-256;5AbXdpz5DcaYXCh9l3eI9ruBosiL5XDU3rxBbBaUO70?module=RA

d. Module Description

There are mainly two module communications to access data byte level and RDF communication charts, for two kinds of relations, version X starting accessible rendition, and describes identifier modules FA and RA. FA module describes SHA-256 substance with different byte information represents with Base64 documentation and also presents 52 character subsequent makes some part have hash URI. RA module describes RDF content covered with different names contain self references and handle hash URI previously generated. To evaluate RDF artifacts which are arrange with serial numbers and finally SHA-256 is similar to FA module specification in data arrangement. For different EDF representations URI generations with different functions shown in table 1

module	function	format	Java	Perl	Python
(general)	RunBatch		✓	✓	✓
File	CheckFile		✓	✓	✓
	ProcessFile		✓	✓	✓
RDF	CheckFile	RDF/XML	✓	✓	✓
		Turtle	✓	✓	✓
		N-Triples	✓	✓	✓
		TriX	✓	–	✓
		TriG	✓	✓	–
		N-Quads	✓	✓	✓
	CheckLargeRdf	*(all of the above)*	✓		
	CheckSortedRdf	*(all of the above)*	✓		
	TransformRdf	RDF/XML	✓		✓
		Turtle	✓		✓
		N-Triples	✓		✓
		TriX	✓	–	✓
		TriG	✓	–	–
		N-Quads	✓		✓
	TransformLargeRdf	*(all of the above)*	✓		
	TransformNanopub	TriX	✓	–	
		TriG	✓	–	–
		N-Quads	✓		
	CheckNanopubViaSparql		✓		

Table 1 Different Types of URI library comparisons used in RDF representations.

e. Implementation Procedure

Based on hash index based nanopublications, our proposed implementation consists following steps:

FileCheck describes file information with individual hash labels

FileProcess defines figures with hash used module (FA) and also describe/modify hash URI if any re-names appear

RDFTransform describes the RDF file info with URI and define the changes file description using RA in design

LargRDFTransform is equal to satisfactory files using whole memory present in stack

NanoPubTransform calls RDFTransform to transform nano-publication files

LargeRDFCheck checks either RDF file using the module RA without any memory utilization specification for transitory files.

Sorted RDFCheck checks arrangement of RDF present entirely or not, also check presents usage of RDF implementation with transitory files

ViaSparqlNanoPubCheck check endpoint of SPARQL & also describes hash URI speak to nano-publication and retrieve nano-publication from the data storage and evaluates to approve it.

BatchRun describes the file directions and execute them in similar manner.

III. EVALUATION OF EXPERIMENTS

This section presents the basic experiments relate to construction of Hash based digital artifacts for different types of semantic web related application. To develop this application, use JAVA latest version with NETBEANS tool latest version used on hash URI based on implementation procedure and RDF files which consist assert provenance and support.

a. Input Data: Data used in our approach is dynamically generates based on user registration, login. Users publish article using their assert, provenance and support data taken from semantic web data sources related to different domains such as real time network applications. Our approach use HASH-256 and base value generates index based Hash value on URI for each user published article with different notations. Basic design for users interface shown in figure 4 with generated URLs.

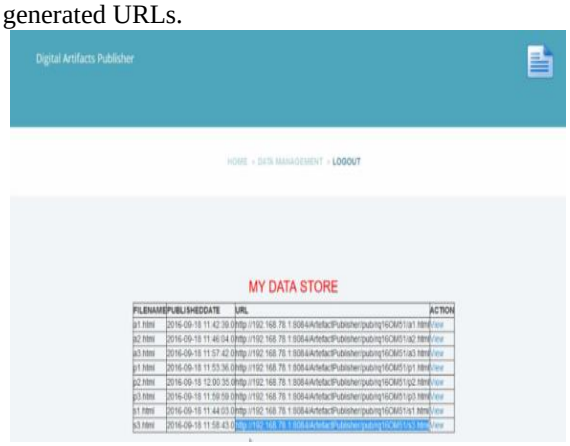

Fig 4 generates different URIs with hash indexes.

b. Results

The tests above spread without a doubt, little RDF files, yet our methodology ought to likewise work for bigger files. Consequently, we played out a second assessment on Bio2RDF, which incorporates a lot bigger files. Main description of this data set contain 945 RDF files in N-dimensional architectural implementation, however some of them mainly lead with well formed errors are stored with stack library (this is library changed based on RDF standard version), these 938 files communicate from 1.5 kB to 100 GB.

We tested on different RDF URI's with advanced hash value generation for trusty URIs and De-centralized Hash URI approach (DHURI) in terms of generation URI with different specifications on semantic web data. Different time comparisons for processing different RDF files shown in figure 5.

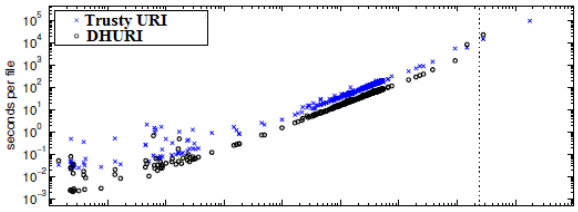

Fig 5 Performance of time with processing RDF files in semantic web application.

Time comparison for hash based digital artifact generation for different web files with respect to assert, provenance and support content of different RDF files generated from semantic web.

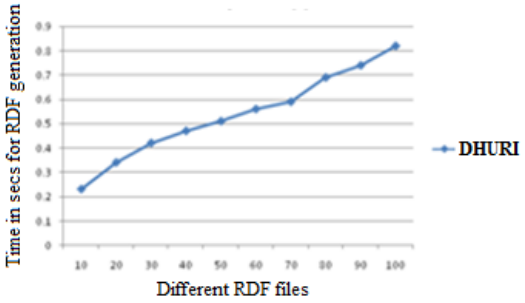

Fig 6 Performance of proposed approach with respect to RDF.

As shown in above figures, our proposed approach gives better results with generation of RDF nanopublications in semantic web related data. Our proposed approach gets verifiable and immutable to share RDF data via semantic data.

IV. CONCLUSION

In this paper, we present De-centralized Hash URI approach (DHURI) to solve the immutable concept of digital artifact generation in nano-publication with unique hash based cryptography value. We discuss proposed approach with efficient nano publication to describe digital artifacts based on assert, support and provenance on web immutable, permanent and verifiable. Together with proposed approaches like nano publications, hash URI contribute and describe indexing on web. Improve the efficiency of web resources and it is important for scientific data, verifiable is appeared for index based hash web data. Further improvement of digital artifact implementation for nano publications to support for different types of file formats i.e CSV, XML, SVG in semantic web data.

REFERENCES

1. Tobias Kuhn, Michel Dumontier, "Hash-URIs for Veri_able, Immutable, and Permanent Digital Artifacts" arXiv:1401.5775v1 [cs.CR] 16 Jan 2014.
2. Tobias Kuhn and Michel Dumontier, "Making Digital Artifacts on the Web Verifiable and Reliable" IEEE TRANSACTIONS ON KNOWLEDGE AND DATA ENGINEERING Vol NO 99 YEAR 2015.
3. Alison Callahan, Jos_e Cruz-Toledo, Peter Ansell, and Michel Dumontier. Bio2RDF release 2: Improved coverage, interoperability and provenance of life science linked data. In Proceedings of ESWC 2013, pages 200{212. Springer, 2013.

4. Alison Callahan, Jos_e Cruz-Toledo, and Michel Dumontier. Ontology-based querying with Bio2RDF's linked open data. Journal of Biomedical Semantics, 4(Suppl1):S1, 2013.
5. Tobias Kuhn. A survey and classi_cation of controlled natural languages. Computational Linguistics, 2014.
6. Tobias Kuhn, Paolo Emilio Barbano, Mate Levente Nagy, and Michael Krauthammer. Broadening the scope of nanopublications. In Proceedings of ESWC 2013,volume 7882 of LNCS, pages 487{501. Springer, 2013.
7. James P McCusker, Timothy Lebo, Cynthia Chang, Deborah L McGuinness, and Paulo Pinheiro da Silva. Parallel identities for managing open government data. IEEE Intelligent Systems, 27(3):55, 2012.
8. James P McCusker, Timothy Lebo, Alvaro Graves, Dominic Difranzo, Paulo Pinheiro, and Deborah L McGuinness. Functional requirements for information resource provenance on the web. In Provenance and Annotation of Data and Pro- cesses, pages 52{66. Springer, 2012.
9. T. Kuhn and M. Dumontier, "Trusty URIs: Verifiable, immutable, and permanent digital artifacts for linked data," in Proceedings of the 11th Extended Semantic Web Conference (ESWC 2014), ser. Lecture Notes in Computer Science. Springer, 2014
10. S. Farrell, D. Kutscher, C. Dannewitz, B. Ohlman, A. Keranen, and P. Hallam-Baker, "Naming things with hashes," Internet Engineering Task Force (IETF), Standards Track RFC 6920, April 2013.
11. E. H¨ofig and I. Schieferdecker, "Hashing of rdf graphs and a solution to the blank node problem," in 10th InternationalWorkshop on Uncertainty Reasoning for the Semantic Web (URSW 2014), 2014,p. 55.
12. H. Van de Sompel, R. Sanderson, H. Shankar, and M. Klein, "Persistent identifiers for scholarly assets and the web: The need for an unambiguous mapping," International Journal of Digital Curation, vol. 9, no. 1, pp. 331–342, 2014.
13. J. McCusker, T. Lebo, C. Chang, D. McGuinness, and P. da Silva, "Parallel identities for managing open government data," IEEE Intelligent Systems, vol. 27, no. 3, p. 55, 2012.
14. J. McCusker, T. Lebo, A. Graves, D. Difranzo, P. Pinheiro, and D. McGuinness, "Functional requirements for information resource provenance on the web," in Provenance and Annotation of Data and Processes. Springer, 2012, pp. 52–66.
15. T. Kuhn, P. Barbano, M. Nagy, and M. Krauthammer, "Broadening the scope of nanopublications," in Proceedings of ESWC 2013, ser. LNCS, vol. 7882. Springer, 2013, pp. 487–501.
16. A. Callahan, J. Cruz-Toledo, and M. Dumontier, "Ontology-based querying with Bio2RDF's linked open data," Journal of Biomedical Semantics, vol. 4, no. Suppl 1, p. S1, 2013.
17. T. Kuhn, C. Chichester, M. Dumontier, and M. Krauthammer "Publishing without publishers: a decentralized approach to dissemination, retrieval, and archiving of data," arXiv preprint arXiv:1411.2749, 2014.
18. Kuhn et al. (2016), Decentralized provenance-aware publishing with nanopublications. PeerJ Comput. Sci. 2:e78; DOI 10.7717/peerj-cs.78.
19. Patrinos GP, Cooper DN, Van Mulligen E, Gkantouna V, Tzimas G, Tatum Z, Schultes E, Roos M, Mons B. 2012. Microattribution and nanopublication as means to incentivize the placement of human genome variation data into the public domain. Human Mutation 33(11):1503_1512 DOI 10.1002/humu.22144.
20. Proell S, Rauber A. 2014. A scalable framework for dynamic data citation of arbitrary structured data. In: 3rd international conference on data management technologies and applications (DATA2014).
21. Sequeda JF, Arenas M, Miranker DP. 2012. On directly mapping relational databases to RDF and OWL. In: Proceedings of the 21st international conference on World Wide Web. ACM, 649_658.
22. Speicher S, Arwe J, Malhotra A. 2015. Linked data platform 1.0. Recommendation W3C.Available at https:// www.w3.org/TR/ldp/ .
23. Verborgh R, Hartig O, De Meester B, Haesendonck G, De Vocht L, Vander Sande M, Cyganiak R, Colpaert P, Mannens E, Van de Walle R. 2014. Querying datasets on the web with high availability. In: Mika P, Tudorache T, Bernstein A, Welty C, Knoblock C, Vrande£i¢ D, Groth P, Noy N, Janowicz K, Goble C, eds. Proceedings of the 13th international semantic web conference. Lecture notes in computer science, vol. 8796. Berlin Heidelberg: Springer, 180_196.
24. Williams AJ, Harland L, Groth P, Pettifer S, Chichester C, Willighagen EL, Evelo CT, Blomberg N, Ecker G, Goble C, Mons B. 2012. Open PHACTS: semantic interoperability for drug discovery. Drug Discovery Today 17(21):1188_1198 DOI 10.1016/j.drudis.2012.05.016.