ISLAND: An Interlinked Semantically-enriched Blockchain Data Framework

Alexandros Kalafatelis[1], Konstantinos Panagos[2], Anastasios E. Giannopoulos[2], Sotirios T. Spantideas[2], Nikolaos C. Kapsalis[2], Marios Touloupou[3] Evgenia Kapassa[3] Leonidas Katelaris[3] Panayiotis Christodoulou[3], Klitos Christodoulou[3], and Panagiotis Trakadas[1]

[1] Department of Port Management and Shipping, National and Kapodistrian University of Athens, Sterea Ellada, 34400 Dirfies Messapies, Greece
kalafatelis.alexander@gmail.com, ptrakadas@uoa.gr
[2] Four Dot Infinity, Athens, Greece research@fourdotinfinity.com
[3] Red Mullet Tech, Paphos, Cyprus
{touloupos, evgeniakapassa,lkateralis, Panchristodoulou87, panklitos}@gmail.com

Abstract. Blockchain is currently one of the most popular technologies, providing privacy, transparency and trust. However, until now, it does not take into consideration the large amount of existing data and standards for decentralized data distribution and processing on the Web, that would leash new opportunities and business innovations for this emerging technology. Moreover, according to the vision of the Semantic Web, a key concept lies on data (semantic) annotations, querying and interlinking. Nevertheless, exporting knowledge resulting from different and possible interlinked blockchain networks, is still a major challenge. To address the aforementioned challenges, we propose ISLAND, a modular framework that is set to expose a unified abstraction layer to any data consumer that aims to infer meaningful knowledge from blockchain generated data, while at the same time enabling the semantic interoperability of them. In addition, a smart manufacturing use case scenario is presented as well as the potential business impacts are discussed.

Keywords: Semantic Blockchain · Linked Data · Machine Learning.

1 Introduction

Blockchain is one of the most innovative and popular technologies nowadays. After it was first introduced by Satoshi Nakamoto, it became known from the rise of cryptocurrencies [1], but soon the scientific community realized its value and started applying it in other fields, such as manufacturing[2], energy [3] and more [4], taking also care in security and privacy issues [5]. In particular, blockchain refers to a data structure, where various forms of data and smart contracts are stored, which are classified into a list of blocks [6, 7]. It is a decentralized network, managed by a peer-to-peer (P2P) network that complies with a protocol for communication between the nodes [8].

Moreover, blockchain environment generates large volumes of distributed data that are publicly available due to their transparent and immutable properties. Even though blockchains are considered a fundamental building block in Web 3.0, they lack software standards that can lead to the emergence of a global, open, interoperable, and semantically rich data-space [9]. The creation of such an interoperable internet and knowledge exchange has been envisioned by Web 3.0, however, it does not yet have the appropriate tools for its pragmatic implementation and development [7, 10].

Moreover, according to the current state-of-the-art, exporting knowledge resulting from different blockchain networks, is a major concern. In particular, the latter is not yet feasible since several challenges are present, such as : i) processing the data, ii) querying the data, (iii) inter-linking data; (iv) integrating the data under a uniform data model; (v) opening the data; (vi) annotating the data and finally in (vii) extracting knowledge [4, 10].

In this paper, we propose the ISLAND framework which aims to address major challenges such as the ones described above. The contributions of this paper include: i) the description of the proposed interlinked and interoperable semantically enriched blockchain data framework, ii) a smart manufacturing use case scenario, of how such a framework can be leveraged in complex real-world systems and iii) potential business impacts.

It is highlighted that the ISLAND architecture has been proposed and selected for funding during the open call of the European Commission project ONTOCHAIN, under the Horizon 2020 framework [11].

The paper is structured as follows: Section 2 gives an overview of the current state-of-the-art and how the ISLAND framework contributes beyond that. Section 3 describes the high-level architecture and the main functionalities of the ISLAND framework. Section 4 presents a use case scenario , while potential business impacts are also identified and discussed in Section 5. Finally, Section 6 summarizes and concludes the paper.

2 State of the Art and Beyond

Blockchain technology provides a decentralized architecture with privacy, transparency and trust but it should also take into consideration the large amount of existing data and standards for decentralized data distribution and processing on the Web. A key concept design of the Semantic Web vision is the semantic annotation of the data and that are easily queried and interlinked. According to Mikroyannidis et al., a Semantic Blockchain, which promotes interoperability between Blockchain networks and the Semantic Web, is needed to get the best out of both technologies [9]. Furthermore, Semantic Blockchain enables the mapping of smart contracts on the blockchain to contextual data about the corresponding data [9].

Towards that direction, Ugarte and Boris introduced a Blockchain Ontology with Dynamic Extensibility [12]. BLONDiE is an OWL ontology for defining the native structure of the blockchain and related data. It covers the two most currently used blockchain protocols (Bitcoin and Ethereum). An example of its abilities is that it is capable of connecting an individual to account data from Bitcoin and Ethereum. Moreover, English et al., demonstrated how blockchain technology can lead to the creation of a more stable Semantic Web, while also a context in which the Semantic Web is utilized to improve blockchain technology itself [13]. Furthermore, Baqa et al., highlighted the fact that although Smart Contracts (SC) are open to the public, it is challenging to discover and utilize such SCs for a wide range of usages since they are compiled in the form of byte-codes without any associated metadata [14]. The latter have motivated the authors to propose the idea of Semantic SC (SSC), a solution that incorporates semantic Web technologies in SCs, which are deployed on the Ethereum Blockchain network, for indexing, searching, and annotating the deployed SCs.

To address the aforementioned challenges and extend the current state-of-the-art, ISLAND focuses on enabling semantic interoperability services for the blockchain technology and beyond. More specifically, ISLAND capitalizes and enhances software blocks brought by enterprises, as background knowledge, to provide a set of tools that would facilitate users to extract pieces of unstructured data from blockchain networks, annotate them with semantic knowledge from Ontologies, and make them interoperable with the use of graph-based formats. Additionally, an innovative idea of the ISLAND framework, lies also in the introduction of the novel Reinforcement Learning algorithm which supports knowledge extraction from large graph sources, based on user requests.

3 Proposed Architecture

ISLAND's high level architecture, which is depicted in Fig.1, consists of several components. Specifically, it incorporates the crawling and indexing services, the Semantic Annotation Module, the AI Knowledge Extraction Module and the Storage and Retrieval services. These components represent stages of heterogeneous data in the whole blockchain's lifecycle. Although they may overlap, they typically happen at different times and are performed by different actors/roles.

Raw data from various data sources are monitored by ISLAND's indexing and crawler services. Such data sources will be: (a) on-chain data (emit events from smart contracts or block data), (b) external data streams (IoT devices) and (c) external linked data (labelled property graphs). However, because of the expected heterogeneity of the data sources, an extra layer which would act as the abstraction layer between those data sources and the ISLAND network is necessary. This component is depicted and referenced as Connectivity Manager in Fig.1. The objective of the Connectivity Manager is to streamline the APIs of the heterogeneous data sources into one API for the Indexing Service to use. This would allow the ISLAND Framework to be agnostic to the wide variety of data sources APIs.

The raw data from various data sources are streamed into the Semantic Annotation Module. These streams are likely to be poor in terms of their semantic annotations and data structure. Depending on the instantiation of the module,

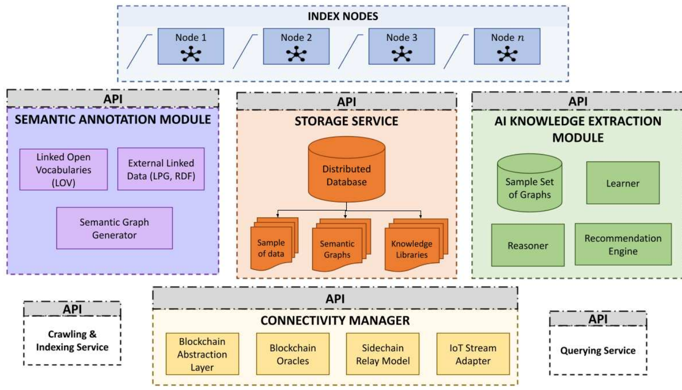

Fig. 1. A high level view of the proposed architecture.

and the domain in hand, a Human Annotator "cherry-picks" terms from various vocabularies with different granularity or expressiveness to semantically annotate a sample of the data. In doing so, human annotations from the sample data are used to train a machine learning engine that can recommend future semantic annotations on similar data. At the same time, Human annotators can point ISLAND's Semantic Annotation Module to a set of open knowledge graphs for inferring semantic terms on entities and their attributes that potentially match to their domain of interest. By observing such annotations, the tool will improve its training thus recommending relevant semantic terms that are likely to be closer to the requirements of the user. The result of this process will be the generation of a final graph dataset that is semantically annotated, but not interlinked with external knowledge graphs and stored within the ISLAND's Distributed Database.

The ISLAND AI Knowledge Extraction Module (ISLAND AI KEM) uses as input the resulting graphs of the Semantic Annotation Module. Moreover, the objective of the AI KEM is to learn over reasoned, contextual knowledge. To achieve this goal, the AI KEM focuses on reasoning over the semantically annotated knowledge graphs using Reinforcement Learning (RL) and then, all the knowledge libraries and the outcome of these components are stored in the Distributed Databases.

We propose the use of RL, because in the first steps of the ISLAND framework, there will be no abundance of data. Moreover, contrary to current publications, we propose on conducting explicit multi-step path reasoning with knowledge of the decision-making stage. Furthermore, this approach combines recommendations with explainability, as it enables the provision of path data in a knowledge graph. The proposed method solves the great issue that most artificial intelligence algorithms face, which is the problem of explainability [15]. Additionally, because it is predicted that the size of the action area is going to be great in size for some nodes, we propose on using a user-conditional action pruning strategy to reduce this size. In several cases, pruning has shown to increase accuracy, a phenomenon which has been observed in a great number of publications [16, 17]. The selected paths are then sorted according to the reward value and are offered to the user.

Through the Indexing Service, the ISLAND platform is able to “scan” through the unstructured raw data and identify any relationships to better index these records for end-users’ benefit. The retrieval of information is accomplished by physically matching keywords in documents with those in a query. Because there are frequently several methods to convey a particular notion, the literal phrases in a user’s query may differ from those in a related resource [18]. ISLAND Indexing Service would allow users to retrieve information on the basis of a topic or meaning of the data. The Indexing Service will allow to quickly retrieve records from the Storage Service.

The Storage Service of the ISLAND framework would use a stateless semantic graph database for storing the final produced graph. The final generated graphs should be stored and be available for future use either from the AI KEM or by the Querying Node in order to reply to an outside consumer request. Moreover, ISLAND’s Storage service would be also responsible to store sample data (Graphs/Unstructured Data, etc.) in order to be used for learning processes from the AI Learner. Finally, knowledge libraries should also be generated storing the knowledge of the AI Recommendation Engine. Thus, on a future consumer request, the knowledge libraries should be searched in case of an already existing reply.

On top of the Storage Service lies the Querying Service, which provides a seamless way for executing queries over a distributed dataset held by different datastores that provide different interfaces. The connectivity for information retrieval can be facilitated via the use of SPARQL.

4 Smart Manufacturing Use Case Scenario

Nowadays, supply chain environments use IoT devices extensively to monitor the chain. However, in spite of the great variety of industries that they can be applied to, they come with certain challenges, due to their centralized structure [19]. Furthermore, because smart-manufacturing services need to exchange machine-readable properties during their end-to-end life-cycle, a semantic system is necessary.

Logistics interactions are supported by the use of electronic message exchanges, such as RosettaNet PIPs (Partner Interchange Processes) which uses XML Schema technologies, in which applications of Semantic Web technology have been repeatedly proposed [20].

This scenario assumes that a company has incorporated IoT (e.g. cameras, thermostats, etc.) devices to monitor its supply chain. In that case, while IoT manufacturing devices generate unstructured raw data, these will be imported in the ISLAND framework, using the crawling service, collecting data based on selected parameters defined by the Curator. Following, the Indexing Service would scan through the unstructured data to identify any relationships between terms collected and their context, to better index these records. The stored data should be provided to ISLAND Semantic Annotation Module, where using the tools included in this module, events are annotated using previously annotated data coming from the blockchain. Then, ISLAND's AI module is triggered by the new sample of data and is trained for future annotations. In the end, semantically enriched graph representations of the data are extracted.

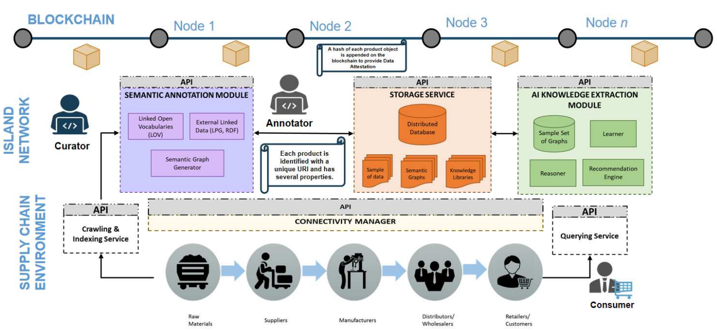

Fig. 2. Detailed view of the ISLAND framework applied in the use case scenario.

5 Potential Business Impacts

The first business impact of the proposed approach is on the smart manufacturing domain. Namely, through the proposed framework, companies will be capable of having traceability and analytics for key logistics operations.

According to Gartner, semantic knowledge graphs and ML algorithms are necessary for Supply Chain operations. It is also predicted that by 2024, half of the global enterprises will have invested in applications supporting artificial intelligence and analytics [21]. Additionally, blockchain technology is getting a great demand. One way smart contracts can be used in Supply Chain, is for generating an invoice when a product delivery reaches the destination, thus resulting in an improved and faster process. The ISLAND framework touches on the heart of the problem and the evolution of the supply chain.

In addition, ISLAND aims to create a new real-world economy thus incentivizing users to participate in ISLAND'S ecosystem to ensure the economic security of the ISLAND ecosystem and the integrity of data being queried is a matter of significant importance. To achieve this, we propose the Island Token (ISL), the native token of ISLAND which will be used by all stakeholders, who participate in ISLAND. ISL is envisioned to be the catalyst for diverse new communities to access enriched semantic graph-based data benefiting real use-cases.

6 Conclusions

This paper proposes ISLAND, an interlinked and interoperable semantically-enriched blockchain data framework, that extends the current state-of-the-art, in the context of interoperability and semantically enriched data coming from blockchain networks. Moreover, the ISLAND framework exposes a human-centric interface to data consumers by generating semantically annotated graphs and enabling knowledge extraction operations on the data.

Furthermore, ISLAND offers numerous tangible advantages stemming from this innovative implementation, ranging from data and metadata validation, to straightforward interoperability, and the use of Ontologies and RDF graphs to extract knowledge over different blockchains combined with content external to the blockchain/DLT ecosystem. Additionally, the possible business impacts of such a solution as ISLAND, will offer a unique proposition of Semantic Blockchain with AI, currently in need for various industries and particularly supply chain.

In conclusion, the ISLAND framework envisions a layer of intermediation between the exposed APIs from the participating smart-contract-users and the data consumers. The framework aims to infer meaningful knowledge from smart contracts, while at the same time enabling the semantic interoperability of the data. Finally, this framework offers a fertile layer for marketable interoperable solutions for domains such as healthcare, economy, public services, energy and sustainability, media, entertainment and Industry 4.0.

Finally, the next step in our work is to complete the design and implementation of all the components of the proposed framework, while we aim to streamline and enhance the interfaces between the different components and the entities. What is more, we plan to implement a Proof of Concept which will be evaluated firstly against the presented manufacturing use case. Additionally, the authors envision to further extend and evaluate the presented approach and multiple heterogeneous use cases, towards the creation of a semantically enriched blockchain universe.

References

1. Nakamoto, S.: Bitcoin: A peer-to-peer electronic cash system. Decentralized Business Review. 21260 (2008)
2. Lagutin, D., Bellesini, F., Bragatto, T., Cavadenti, A., Croce, V., Kortesniemi, Y., Leligou, H., Oikonomidis, Y., Polyzos, G., Raveduto, G., Santori, F., Trakadas, P., Verber, M.: Secure Open Federation of IoT Platforms Through Interledger Technologies - The SOFIE Approach. 2019 European Conference on Networks and Communications (EuCNC). 518-522 (2019)
3. Kapassa, E., Themistocleous, M., Quintanilla, J., Touloupos, M., Papadaki, M.: Blockchain in Smart Energy Grids: A Market Analysis. Information Systems **402**, 113-124 (2020)
4. Makridakis, S., Christodoulou, K.: Blockchain: Current Challenges and Future Prospects/Applications. Future Internet **11**, 258 (2019)

5. Lagutin, D., Anton, P., Bellesini, F., Bragatto, T., Cavadenti, A., Croce, V., Fotiou, N., Haavala, M., Kortesmiemi, Y., Leligou, H., Manzoor, A., Oikonomidis, Y., Polyzos, G., Raveduto, G., Santori, F., Siris, V., Trakadas, P., Verber, M.: The SOFIE Approach to Address the Security and Privacy of the IoT using Interledger Technologies. Security and Privacy in the Internet of Things: Challenges and Solutions **27**, 76-93 (2020)
6. Ali Syed, T., Alzahrani, A., Jan, S., Siddiqui, M., Nadeem, A., Alghamdi, T.: A Comparative Analysis of Blockchain Architecture and its Applications: Problems and Recommendations. IEEE Access **7**, 176838–176869 (2019)
7. Ugarte, H.: A more pragmatic Web 3.0: Linked blockchain data. Bonn, Germany (2017)
8. Wu, H., Cao, J., Yang, Y., Tung, C., Jiang, S., Tang, B., Liu, Y., Wang, X., Deng, Y.: Data Management in Supply Chain Using Blockchain: Challenges and a Case Study. In: 28th International Conference on Computer Communication and Networks (ICCCN), pp. 1–8. (2019)
9. Mikroyannidis, A., Third, A., Domingue, J.: A Case Study on the Decentralisation of Lifelong Learning Using Blockchain Technology. Journal of Interactive Media in Education **2020**(1), 23 (2020)
10. Belchior, R. Vasconcelos, A., Guerreiro, S., Correia, M. A Survey on Blockchain Interoperability: Past, Present, and Future Trends." ArXiv abs/2005.14282 (2020)
11. ONTOCHAIN, https://ontochain.ngi.eu. Last accessed 24 Mar 2021
12. Ugarte-Rojas, H., Boris, C.: BLONDiE: Blockchain Ontology with Dynamic Extensibility. ArXiv abs/2008.09518 (2020)
13. English, M., Auer, S., Domingue, J: Block chain technologies and the semantic web: A framework for symbiotic development. In: Computer Science Conference for University of Bonn Students, pp. 47–61. (2016)
14. Baqa, H., Truong, N., Crespi, N., Lee, G., le Gall, F.: Semantic Smart Contracts for Blockchain-based Services in the Internet of Things. In:2019 IEEE 18th International Symposium on Network Computing and Applications (NCA), pp. 1–5. (2019)
15. Makni, B., Abdelaziz, I., Hendler, J.: Explainable Deep RDFS Reasoner. ArXiv, abs/2002.03514 (2020)
16. Han, S., Pool, J., Tran, J., Dally, W.: Learning both Weights and Connections for Efficient Neural Network. ArXiv, abs/1506.02626 (2015)
17. Suzuki, T., Abe, H., Murata, T., Horiuchi, S., Ito, K., Wachi, T., Hirai, S., Yukishima, M., Nishimura, T.: Spectral-Pruning: Compressing deep neural network via spectral analysis. ArXiv, abs/1808.08558 (2018)
18. Rosario, B.: Latent semantic indexing: An overview. INFOSYS, 240, 1-16 (2000)
19. Atlam, H., Alenezi, A., Alassafi, M., Wills, G.: Blockchain with Internet of Things: Benefits, Challenges, and Future Directions. International Journal of Intelligent Systems and Applications **10**, 40–48 (2018)
20. Cardoso, J., Bussler, C.: Mapping between heterogeneous XML and OWL transaction representations in B2B integration. Data and Knowledge Engineering **70**, 1046–1069 (2011)
21. Gartner Predicts the Future of Supply Chain Technology, https://www.gartner.com/smarterwithgartner/gartner-predicts-the-future-of-supply-chain-technology/. Last accessed 24 Mar 2021