KNOWLEDGE GRAPH BASED SEMANTIC MODELING FOR PROFILING IN INDUSTRY 4.0

Siraj Munir, Syed Imran Jami, Shaukat Wasi

Computer Science Department, Mohammad Ali Jinnah University
e-mails: sirajmunir93@gmail.com , imran.jami@jinnah.edu,
shaukat.wasi@jinnah.edu
Pakistan

Abstract: In this paper we present a framework for profiling of workers and employees in industry setting using the camera feeds. With the incorporation of Internet of Things with Big Data and Semantic Web, we provide fast communication and cooperation among humans in real-time within the industry. It aims at supporting intelligent queries about the workers working in the Industry by using temporal knowledge graph generated from the multi attributed logs identified from the surveillance video. The paper facilitates integration of the profiling knowledge graph with existing industry knowledge graph for unified view of profiling log with worker's information. Decentralized decision making in real time can be made with this system which is one of the core areas of Industry 4.0. Knowledge Graph is proposed as generic for reusability in any industries and enterprise for profiling purpose. The results show the highlights of the queries about working in the industry without the involvement of any resource. The paper is concluded with shortcomings in this work along with the solution

Key words: Industry 4.0, Profiling, Temporal Knowledge Graph, Semantics, Smart Industry

1. INTRODUCTION

Fourth automation based industrial revolution also known as Industry 4.0 was introduced in 2011 by German government. Industry 4.0 is dependent on several areas including Internet of Things (IoT), Cyber Physical Systems (CPS), Smart City, Digital Twins etc. Internet of Things (IoT) is an extension of internet which integrates physical devices with real world objects. Cyber Physical Systems (CPS) are composed of strong interconnection between physical and software components. These systems are multidisciplinary in nature as they include cybernetics, mechatronics, embedded systems etc.[1]. Smart City is a modern city concept that uses different IoT devices for data acquisition and then use this acquisitioned data for resource management. Digital Twin is a digital replica of living or non-living physical entity. By bridging the physical and virtual world, data is transferred seamlessly allowing the virtual entity to exist simultaneously with the physical entity. Semantic and profiling information is key to a successful implementation of Industry 4.0. Due to mass generation of data we need to tackle irrelevant information. To understand importance of data we need semantics. Till date semantic information for profiling data has been partially addressed by some researchers [2-6]. In this work we have proposed a semantic framework which integrates semantic information of Employee/worker through Knowledge Graph, temporal profiling information and facial recognition. Using the proposed framework user can answer semantic queries of workers, working in Smart Industry. For running semantic queries, we have designed a query engine. Some sample semantic queries are as follows:

1. How many workers were working on machine A?
2. At timestamp XX where was Manager A?
3. How can we introduce person A to person B?
4. How many people were present at Canopy location?
5. Which locations did worker A visited today?

This work is further detailed in three sections. The next section discusses the latest trend in Industry 4.0. Section 3 and 4 details proposed architecture and discussion on results. The paper is concluded with the benefits of this architecture with limitations.

2. RELATED WORK

Industry 4.0 or Smart Industry is a growing area which integrates automation with manufacturing. In this section we will review literatures published in last three years. In a work proposing involvement of human in production loop for industry 4.0 [7], it is discussed that humans can help to mitigate issues like machine failure, decision failure and will enhance the performance of production lifecycle. Authors also proposed a novel approach through which they were capable to utilize human's cognitive capabilities along with contextual information management for Industry 4.0. In [8] authors discussed the personal data privacy and its challenges in the context of Industry 4.0. Industry 4.0 enabled system often involve device to device communication. During such communication devices preserve some information. Securing this preserved information is crucial as it can lead to data security attacks. Authors have investigated some possible privacy attacks in Industry 4.0 environment. Authors also highlighted the role of personally identifiable information and contextual privacy awareness in the context of Industry 4.0. Specialized use cases for Unmanned Vehicle in Industry 4.0 were discussed in [9]. Moreover, authors also studied the wireless technologies that can be used for deployment of Unmanned Vehicles for Industry 4.0 scenarios. Authors of literature [10] discussed role of Supply Chain Management Marketing in Industry 4.0. Authors also identified the critical areas which can be used to aid Industry 4.0 oriented Supply Chain Management. The identified areas include Industrial IoT, Cloud Computing, Big Data analytics and customer profiling. Big Data analytics help users to reveal useful patterns from data through analysis. Work in [11] reviewed the Business Intelligence (BI) aspects for Industry 4.0 in which authors discussed the value of Business Intelligence and research trends and gaps. Authors also emphasized on the need of enhanced models for business research in Industry 4.0 environment. CyberFactory#1 model was proposed in [12]. Using the proposed model authors resolved the security concerns that arise in Industry 4.0 based system including avionics quality monitoring. Authors also tested the proposed model on real world operational environment including transportation, automotive, electronics and machine manufacturing. Unmanned Aerial Vehicle (UAV) based system was proposed in [13] which was designed to do autonomous inventory monitoring like item traceability. For tracing the industrial items authors used (Radio Frequency Identification) RFID tags. Moreover, for decentralization authors used Block chain mechanism to retrieve the inventory data collected from UAV. In a survey on business model for Industry 4.0 [11] authors reviewed last five year published articles. Authors reviewed several business models for Industry 4.0 that were mainly ubiquitous in nature. Here ubiquity expresses relationship between customer and supplier. Knowledge Graph based embedding approach was proposed in [14] in which authors model embedding to convert graph into vector space so that it can be trained using traditional machine learning approaches. Authors also discussed proposed model and its implication for Industry 4.0 scenarios. In [15] authors discussed latest techniques for modelling knowledge in Knowledge Graph. Proposed literature also shed lights on creation of Knowledge Graphs by using unstructured data sources. Further authors experimented proposed approach on cyber physical systems. Authors in [16] proposed a model which integrates the solution to workers in Industry 4.0. Proposed model only targeted three areas (i) Data Organization, (ii) Content extraction and filtering and (iii) Support to user/worker during use and maintenance of machine. Proposed model was designed to assist workers to resolve machine failures. In recent years several approaches have been proposed for data acquisition like in [17]. Image segmentation technique integrated with wearables devices were used for data acquisition. Authors validated the proposed approach with four data transformation approaches including (i) Raw plots, (ii) Multichannel plots, (iii) Spectrogram and (iv) Spectrogram and Shallow features. For validation authors trialed the proposed approaches with three public datasets. Big data processing framework for Industry 4.0 scenarios was proposed in [18]. Framework was also experimented with big data predictive maintenance task. For experiment authors used heterogeneous sources for data acquisition. Moreover, authors also piloted the proposed framework for analyzing multisource industrial data and predicted life expectancy of machine equipment. In literature [19] authors proposed big data driven predictive schemes for healthcare and Industry 4.0 scenarios. Authors proposed different models for resolving critical issues which may arise in Industry 4.0. These issues include preprocessing methods for industrial big data, associate analysis based feature processing, deep learning based prognostics model, spark platform based parallel computing etc.

In this work we attempted to integrate Big Data for profiling with Knowledge Graph to answer semantic queries in Industry 4.0 settings using IoT devices. We have modelled Workers and Managers to monitor the activities through a semantics enabled data centric approach. For data acquisition of profiling data, we have used facial recognition techniques and smart devices. By using our proposed framework administrator can run semantic queries over Knowledge Graph. In the next section we will detail system architecture with the discussion on data and semantics modeling.

3. SYSTEM ARCHITECTURE

Layered architecture is proposed for gathering of relevant information as logs and extraction of semantic information from logs for answering semantic queries. With the ubiquitous penetration of electronic gadgets including cameras, GPS, E-Tags at intersections, automobiles and IOT based devices for smart homes, lot of data can be gathered and used from these devices after cleaning for the purpose of profiling. The gathered data will be helpful in detecting the activity at different places of industry in a seamless and autonomous way. In this work we used open source CNN architecture for facial recognition [20]. We trained our model on 100 images per person. Using aforementioned CNN setup, we got accuracy of 97%. The subsequent subsections will discuss different components of our profiling architecture.

3.1. Data modelling

Data modelling is a technical process through which we determine the requirements for particular system. Requirements can be conceptual, logical or physical. Figure 1 shows the proposed data model for maintaining Profiling data in the domain of smart industry. The data is extracted from curated data sources including Surveillance Camera and Surveillance devices as extension. Face recognition module is developed to detect the employee which will help in determining his three attributes: {ID, timestamp, location}. This module is developed using Python that detects and transfers the logs (instead of multimedia data) to centralized database thus reducing the network cost.

For data representation, we have used NoSQL due to its schema less architecture which makes it efficient and more effective. For storage purpose, it uses key, value paired data whose structure resembles with attributes of the entity. The data is dumped in centralized MongoDB. The attributes of profiling information are uniform and finite however they increase exponentially with the passage of time due to the movement of workers in industry. Therefore, the logs are dumped in Big Data representation.

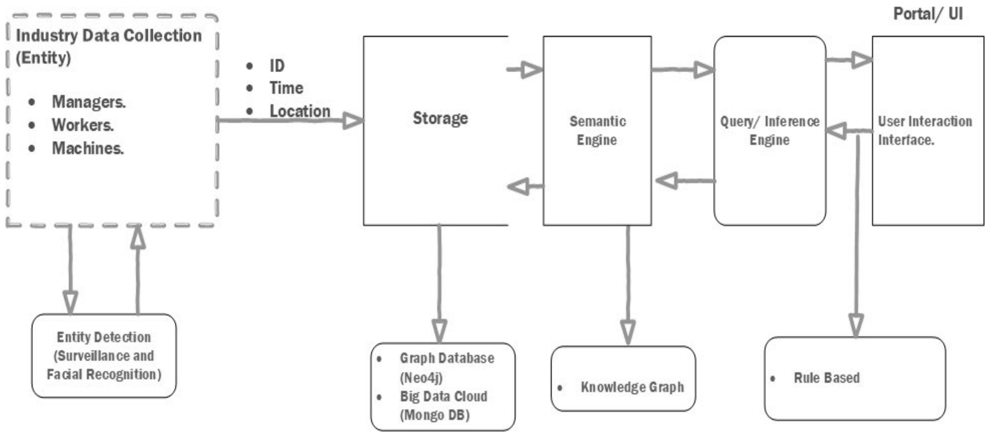

Figure 1. Data model for Surveillance and Sousviellance based employee profiling

The use of NoSQL for Big Data is its extensibility for profiling data that has the properties of 3V's including velocity, variety, and volume and is therefore be implemented in Big Data [21]. For storage we have used two centralized storage tools (i) MongoDB an open source document based NoSQL database and (ii) Neo4j an adjacency based NoSQL graph database. MongoDB is used for maintaining Employees' Profiling Big Data logs and Neo4j has been used for the implementation of Knowledge Graph. Finally, for querying and inferencing we have web portal through which the authorized users (HR, Managers, and Employers) of profiling framework can ask Semantic Questions by querying Knowledge Graph. The same module has the extension for vehicle profiling in industry 4.0 scenario to identify the movement of loading vehicles.

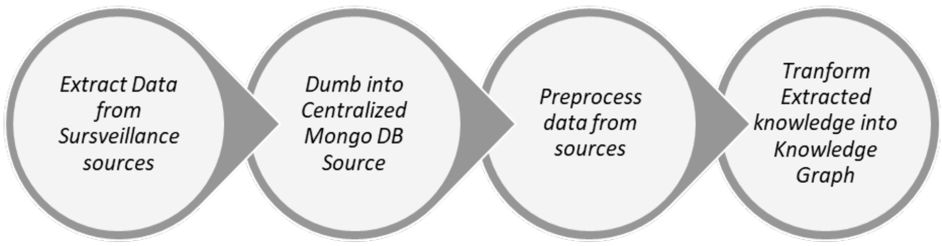

Figure 2. Pipeline for Knowledge Graph management

The pre-processing step in figure 2 requires data cleaning process before extraction of semantic information. This step resolves the redundancy and inconsistency of profiling logs from curated sources. The pre-processed data becomes an input for Semantic model which is represented as Knowledge Graph. The next sub section will discuss the representation of Semantic Information.

3.2. Semantic Modelling

Knowledge Graph for Profiling data needs the integration of temporal information to represent real time information. Maintaining temporal information in graphs is a complex task. The movement of workers and employees is dynamic in nature. To keep the profiling log, we need to gather their data at every instance of time to answer intelligent and temporal queries. In the test case of Industry 4.0 scenario the information of person is updating after every instance periodically. The problem of maintaining temporal information in Knowledge graphs has also been addressed in [21, 22]. Semantic Information can be modelled either in the form of Knowledge Graph [23] or Ontology [24]. Both models follow the structure of triples: Subject, Predicate, Object as shown in Table 1. However, for temporal information we need to add the attribute of Timestamp in the extended triples: Subject, Predicate, Object, Timestamp. Table 1 shows an example of maintaining temporal information in Knowledge Graph which is also proposed by [22]. In this work we implemented this model as shown in Figure 1 for Industry 4.0 scenario.

Table 1. Examples for maintaining Knowledge Graph

Information	**Relationship**
(Siraj, Work_at, XYZ Textiles)	Work_at
(Siraj, Located_at, XYZ Textiles, 2019-02-19 03:53)	(Located_at, Timestamp)

In figure 3 the four tuple attributes are transformed into two integrated triples to represent temporal information. The Employee entity includes workers, operators, managers etc. while location represents the placement of surveillance camera.

Figure 3. Graph model for maintaining Temporal profile

Table 2 shows the snapshot of our Knowledge Graph for Profiling of Industry workers. Several interesting and intelligent queries have been answered in an autonomous way. The next section shows the Results and nature of answered queries.

Table 2: Tabular snapshot of Knowledge Graph

Person	Locations	Relationships	Timestamp
Manager342	Manager Office	Located_at	2019-02-19 03:53
Worker112	Gate1	Located_at	2019-02-19 04:53
Mechanic132	Weaving Hall	Located_at	2019-02-19 05:55
Manager345	HR Department	Located_at	2019-02-19 10:53
Worker112	Weaving Hall	Located_at	2019-02-19 02:53
Worker113	Cafeteria	Located_at	2019-02-19 01:53
Manager342	Canopy	Located_at	2019-02-19 01:50
Worker311	Cafeteria	Located_at	2019-02-19 01:30
Guard112	Manager Office	Located_at	2019-02-19 04:53

4. RESULTS AND DISCUSSION

4.1. Querying profiling data on Knowledge Graph

Our key contribution in this work is the integration of semantic profiling information with Knowledge Graph of the person. The dataset was extracted from NoSQL database having key features (i) Employee ID, (ii) timestamp and (iii) Location. For data acquisition we have used IoT devices including RaspberryPi for processing of pictures and PiCamera for capturing of pictures. For the sake of clarity and easiness the prototype model we considered only three locations for data acquisition. In this section, we will discuss results that were achieved by implementing proposed model on large scale dataset. Figure 4 shows the information generated from an employee's Knowledge Graph generated from existing dataset.

This panel will be visible to Managers and CEO of the industry showing the general information about the worker. He can access his profiling information by clicking on the link. The next subsection details the intelligent queries to which one gets the answer.

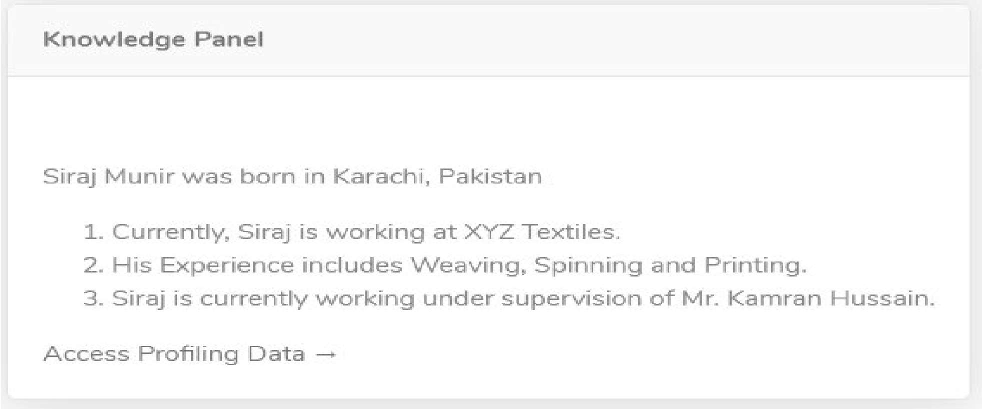

Figure 4. Employee Knowledge Graph

Q1: How many workers were present at Machine A Location today?

This query is translated in Cypher as MATCH (E:Person) WHERE E.Location = “Machine A” RETURN E;

The answer to this query generates following result as figure 5:

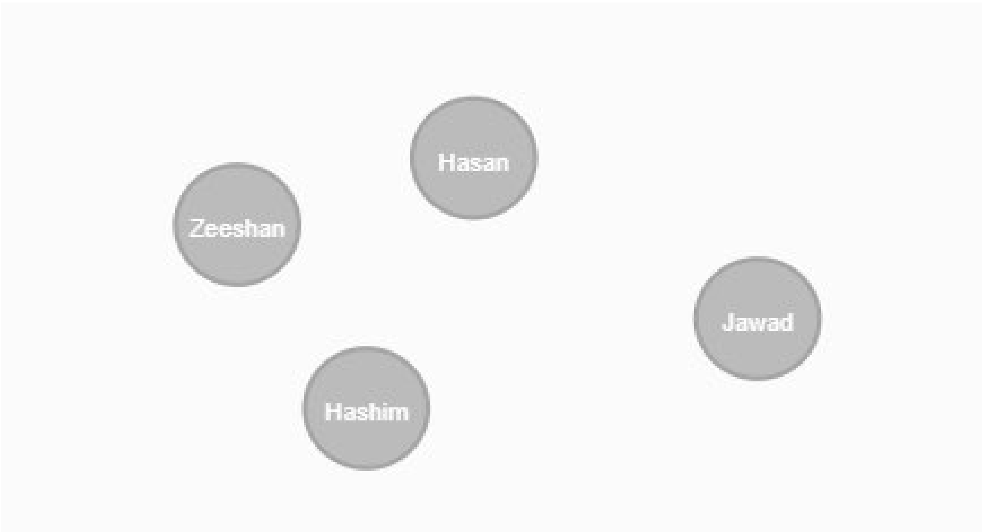

Figure 5. Employees present at Machine A location

Using proposed knowledge graphs we can also visualize complete profiling and semantic information. The following screenshot in figure 6 is generated by Neo4J that shows the results of profiled entities at specific location over different timestamps. The further semantic information can be extracted by zooming the entity as mentioned above.

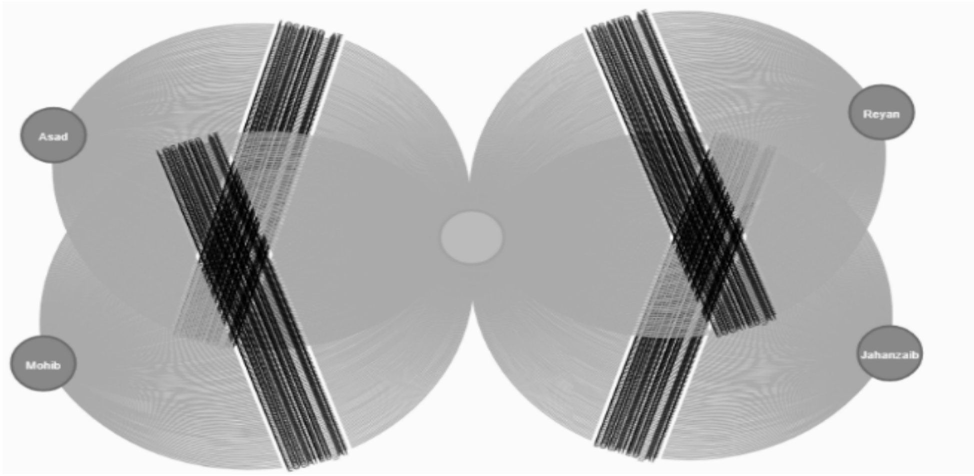

Figure 6. Neo4j Profiling Knowledge Graph

As discussed in previous sub section for semantic querying we have developed a web interface. Using our web interface user can search and visualize the semantic query results. Following screen shot in figure 7 depicts our query engine interface.

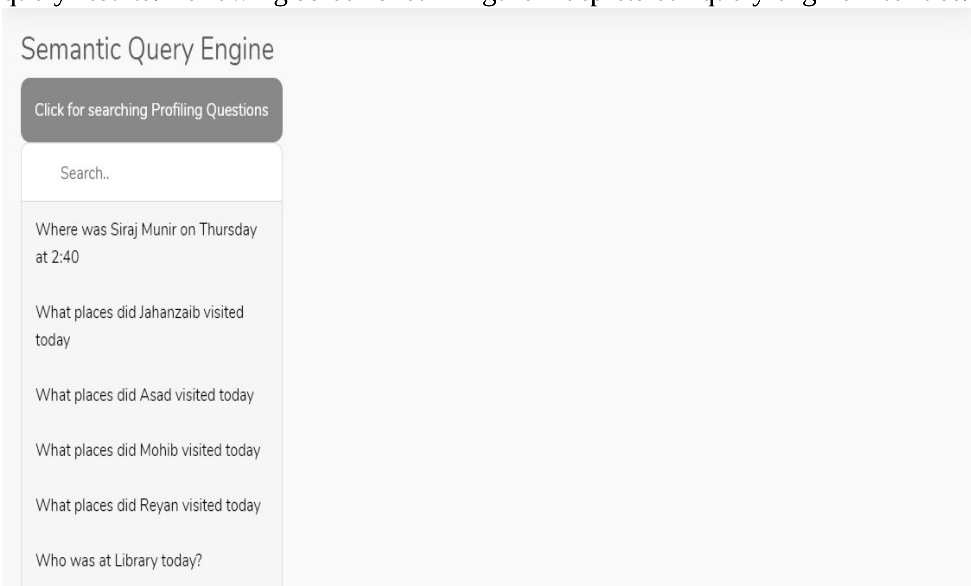

Figure 7. Semantic Query Engine

Q2. Who was the worker X at timestamp XX?

This query is translated in Cypher as MATCH (E:Person) WHERE E.timstamp = “6/3/2019 1:00” AND E.id = “123” RETURN E;

The answer to this query generates following result as figure 8:

DataTable

	Location
1586	Canopy

Figure 8. Query result for finding employee at specific timestamp

Q3. Who visited Canopy today?

This query is translated in Cypher as MATCH (E:Person) WHERE E.Location = “Canopy” AND E.timestamp = “6/3/2019” RETURN E;

The answer to this query generates following result as figure 9:

DataTable

	Name
0	Jahanzaib
1	Reyan
3	Asad
1586	Siraj Munir
1587	Imran Jami
1588	Shaukat Wasi
1597	Anus bin Murtaza
1598	Abdullah Qamar

Figure 9. Query result for location specific search

Q4: Which places did worker X visit today?

This query is translated in Cypher as MATCH p=(Person)-[r:LOCATED_AT]->() WHERE Person.Name = "Siraj Munir" RETURN p

The answer to this query generates following result as figure 10:

DataTable

	Location
0	Canopy
306	Cafe

Figure 10. Query result for worker visited locations

The accuracy of answering to intelligent queries on Knowledge Graph is highly dependent on data gathering as discussed in previous section. The accuracy of capturing employee from captured photographs is highly dependent on camera resolution and image based dataset of workers. At least 100 images of each worker in different pose are needed for existing algorithms of image processing which can further be improved. However, it is considered as out of scope for this work. The image data is transformed into attributes as Big Data due to high velocity of data generation.

4. CONCLUSION

Citizen Profiling is of the recent areas in computing. It requires input from multidisciplinary experts to resolve the issues of privacy, law, and computing. In this work we extended the idea to Industry 4.0 which is considered as the revolution in smart industry. It helps the industry controlling authorities to watch the activities of workers through the use of data and query without any need to physically watch the camera. A related approach is the profiling of social media however research shows that the available information is extremely limited with respect to movement of workers. In this work we proposed a semantic framework which enables to integrate Temporal Knowledge Graph with the profiling data. By employing Web Interface based on HCI, this system will be transformed into the search engine of the workers.

This area leads to several exciting areas in Intelligent Information Systems. Inferencing is an intuitive way to devise the connecting blocks from data that will lead to the extraction of knowledge. This knowledge may result into false facts due to the presence of incomplete (or garbage) information due to the generation of curated data. The inference engine over the knowledge graph will provide the solution for misleading conclusion through wrong interpretations. Therefore integration of inferencing with knowledge graph will solve the problem of semantic failure. Another domain is privacy. This system is currently vulnerable against it due to open access of profiling data. An adaptive system over the inference engine can be incorporated for selective access on the basis of user role to develop context aware interfaces. For efficient data management, a decentralized system for storage can easily be incorporated where storage can be distributed as zones in the industry. This will also partially solve the issue of security. For better domain modeling towards universal adaptability specialized ontology for Industry 4.0 domain integrated with profiling information will help our system to further refine our Knowledge Graph.

REFERENCES

[1] E. A. Lee, "Cyber Physical Systems: Design Challenges," in *2008 11th IEEE International Symposium on Object and Component-Oriented Real-Time Distributed Computing (ISORC)*, 2008, pp. 363-369.

[2] J. W. Ansari, "Semantic Profiling in Data Lake," RWTH Aachen University, Germany, 2018.

[3] C. Segalin, D. S. Cheng, and M. Cristani, "Social profiling through image understanding: Personality inference using convolutional neural networks," *Computer Vision and Image Understanding,* vol. 156, pp. 34-50, 2017/03/01/ 2017.

[4] F. Leal, H. González–Vélez, B. Malheiro, and J. C. Burguillo, "Semantic Profiling and Destination Recommendation based on Crowd-sourced Tourist Reviews," Cham, 2018, pp. 140-147.

[5] A. Mazayev, J. A. Martins, and N. Correia, "Interoperability in IoT Through the Semantic Profiling of Objects," *IEEE Access,* vol. 6, pp. 19379-19385, 2018.

[6] S. Park, A. Matic, K. Garg, and N. Oliver, "When Simpler Data Does Not Imply Less Information: A Study of User Profiling Scenarios With Constrained View of Mobile HTTP(S) Traffic," *ACM Trans. Web,* vol. 12, pp. 1-23, 2018.

[7] C. Emmanouilidis, P. Pistofidis, L. Bertoncelj, V. Katsouros, A. Fournaris, C. Koulamas, *et al.*, "Enabling the human in the loop: Linked data and knowledge in industrial cyber-physical systems," *Annual Reviews in Control,* 2019/03/19/ 2019.

[8] O. Hassan, K. Chul-Soo, and Y. Jinhong, "Personal Data Privacy Challenges of the Fourth Industrial Revolution," in *2019 21st International Conference on Advanced Communication Technology (ICACT),* 2019, pp. 635-638.

[9] F. Foresti and G. Varvakis, "Ubiquity and Industry 4.0," in *Knowledge Management in Digital Change: New Findings and Practical Cases*, K. North, R. Maier, and O. Haas, Eds., ed Cham: Springer International Publishing, 2018, pp. 343-358.

[10] T. M. Fernández-Caramés, O. Blanco-Novoa, M. Suárez-Albela, and P. Fraga-Lamas, "A UAV and Blockchain-Based System for Industry 4.0 Inventory and Traceability Applications," in *Multidisciplinary Digital Publishing Institute Proceedings*, 2018, p. 26.

[11] F.-È. Bordeleau, E. Mosconi, and L. A. Santa-Eulalia, "Business Intelligence in Industry 4.0: State of the art and research opportunities," 2018.

[12] A. Bécue, Y. Fourastier, I. Praça, A. Savarit, C. Baron, B. Gradussofs, *et al.*, "CyberFactory#1 — Securing the industry 4.0 with cyber-ranges and digital twins," in *2018 14th IEEE International Workshop on Factory Communication Systems (WFCS),* 2018, pp. 1-4.

[13] A. Fellan, C. Schellenberger, M. Zimmermann, and H. D. Schotten, "Enabling Communication Technologies for Automated Unmanned Vehicles in

Industry 4.0," in *2018 International Conference on Information and Communication Technology Convergence (ICTC)*, 2018, pp. 171-176.

[14] M. Garofalo, M. A. Pellegrino, A. Altabba, and M. Cochez, "Leveraging Knowledge Graph Embedding Techniques for Industry 4.0 Use Cases," *CoRR,* vol. abs/1808.00434, 2018.

[15] A. Rettinger, S. Zander, M. Acosta, and Y. Sure-Vetter, "Semantic Technologies: Enabler for Knowledge 4.0," in *Knowledge Management in Digital Change: New Findings and Practical Cases*, K. North, R. Maier, and O. Haas, Eds., ed Cham: Springer International Publishing, 2018, pp. 33-49.

[16] G. Lotti, V. Villani, N. Battilani, and C. Fantuzzi, "Towards an integrated approach for supporting the workers in Industry 4.0," in *2018 IEEE Industrial Cyber-Physical Systems (ICPS)*, 2018, pp. 609-614.

[17] X. Zheng, M. Wang, and J. Ordieres-Meré, "Comparison of Data Preprocessing Approaches for Applying Deep Learning to Human Activity Recognition in the Context of Industry 4.0," *Sensors,* vol. 18, p. 2146, 2018.

[18] J. Yan, Y. Meng, L. Lu, and L. Li, "Industrial Big Data in an Industry 4.0 Environment: Challenges, Schemes, and Applications for Predictive Maintenance," *IEEE Access,* vol. 5, pp. 23484-23491, 2017.

[19] J. Yan, Y. Meng, L. Lu, and C. Guo, "Big-data-driven based intelligent prognostics scheme in industry 4.0 environment," in *2017 Prognostics and System Health Management Conference (PHM-Harbin)*, 2017, pp. 1-5.

[20] PyImageSearch. (18-11). [Web Post (Tutorial)]. Available: https://www.pyimagesearch.com/2018/06/18/face-recognition-with-opencv-python-and-deep-learning/

[21] S. Gottschalk and E. Demidova, "EventKG: A Multilingual Event-Centric Temporal Knowledge Graph," presented at the In Proceedings of the 15th Extended Semantic Web Conference (ESWC 2018), 2018.

[22] A. Garcia-Duran, S. Dumancic, and M. Niepert, "Learning Sequence Encoders for Temporal Knowledge Graph Completion," 2018, pp. 4816-4821.

[23] J. o. S. Web. (2015). *JWS Special Issue on Knowledge Graph*. Available: http://www.websemanticsjournal.org/index.php/ps/announcement/view/19

[24] N. F. Noy and D. McGuinness, *Ontology Development 101: A Guide to Creating Your First Ontology* vol. 32, 2001.

Information about the authors:

Dr. Syed Imran Jami is working as Associate Professor in Department of Computer Science, Mohammed Ali Jinnah University, Karachi. His areas of interest include Semantic Web, Social Computing, and Software Agents.

Dr. Shaukat Wasi is working as Associate Professor in Department of Computer Science, Mohammed Ali Jinnah University, Karachi. His areas of interest include Information Extraction, HCI, and Knowledge Management.

Siraj Munir is the Graduate Student at Mohammed Ali Jinnah University, Karachi. He is also working as Research Assistant working in the areas of Semantic Web and Machine Learning.

Manuscript received on 02 December 2019