2nd International Conference on Current Research Trends in Engineering and Technology
© 2018 IJSRSET | Volume 4 | Issue 5 | Print ISSN: 2395-1990 | Online ISSN : 2394-4099
Themed Section: Engineering and Technology

A Review - From Data to Actionable Knowledge

Pinkal Shah, Trilok Suthar, Digvijaysinh Mahida, Shivam Upadhyay
Assistant Professor, IT Department, Sigma Institute of engineering, Vadodara, Gujarat, India

ABSTRACT

Over the years, Speed of data generation has become very frequent and advanced technologies have provided more facilities to generate and gather data. These data are continuous and associated with many other things like decision making and data stream mining. Among multiple technologies Internet of Things (IoT) plays significant role in generating different types of data and make computational of data more challenging. In this paper, Extending the current technology providing connection and internetworking between devices and physical objects or "things," is a growing trend that's often referred to as the Internet of Things (IoT). Sharing real-world data into the Web, with its large sources of data, and providing Web-based interactions between humans and IoT object is what the Web of Things (WoT) stands for. Here, the Big Data issues in the WoT, discuss the challenges of extracting actionable data and insights from raw object data.

Keywords: Internet of Things, Data Analysis, Web of Things, Big Data

I. INTRODUCTION

The speed of data produced and communicated over and between objects of IoT and the Web is quickly increasing. Produced data is in structured, unstructured and semistructured form. This includes textual content, multimedia content from social networking sites or sensible objects. Because of IoT rapidly growing data is related to physical observations, measurements from real world. The evolution of physical world data collection is from sensing devices like sensors and actuators, modems, smartphones or network connected devices. This trend is accelerated, and it will keep on generating data.

Extending the current technology providing connection and internetworking between devices and physical objects or "things," is a growing trend that's often referred to as the Internet of Things (IoT). Another related term to IoT is Internet of Everything (IOE), which generalized the role of human sensing, via social media, to accompaniment physical sensing implied by IoT. Integrating real world data to web and providing interactions with resources or objects is referred as Web of Things (WoT). Collected data is always in different types like light, voice and video [1]. WoT resources can be universal and can be controlled by power, storage, computing, and connectivity. The nature of devices and large range of data make understanding, integrating and processing the real-world data on web a tough task but web data is not limited to only sensors. Integration of real-world, social media and cyber data enables resources to develop application and services to make smart decision [2,3]. WoT is also as large as IoT in terms of size and volume. Objects that produce data are working in dynamic and unpredictable environment.

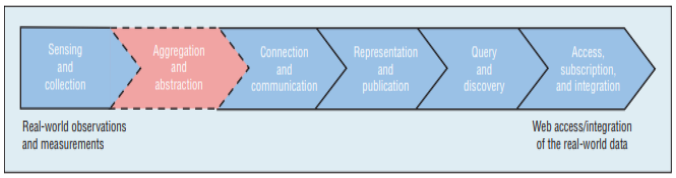

Figure 1: A data access and production.

The dynamic environment and data sources make efficient use of the WoT data on a global scale a challenging task. Figure 1 shows the data access and production chain of physical world. Data generated and stored by machines, sensors or any other smart device. The data can be transform, load with high level abstraction. The data is loaded and stored with metadata for extraction or representation of data. Data aggregation also happens at last stage while combining from various resources and different types. Aggregation also recognize the importance and role of data created from different kind of raw data. Most of the data is outcome of traffic prediction, healthcare and Industrial IoT.

II. DATA PRODUCTION AND ACCESS CHAIN

A. Sensing and Data Collection

Sensor devices, human sensing resource and social media are main sources of generating and storing real-world data which can be integrated, transformed and accessed over web. Main sources generated data on high scale, but they are not stable in terms of quality. Quality of data vary over intervals. It is challenging in Big data to shape quality, validation and trust on collected data from real-world. In most of the cases IoT is dynamic with connected devices so source of data will be unknown sometimes. A real-world data is related to environment, human so privacy and trust on data will always remain a key issue. WoT data is accessible to many so processing by different accessible individuals will create more challenging task to manage security of data. For example, in a smart traffic environment where the sensing devices collect data related to vehicle location and its condition. This data will be access by multiple agencies and types of use of this data can raise significant security and privacy concerns.

B. Connection and Communication

The WoT resources, runs on power and have processing limit but this device have multiple interfaces to connect with wireless connections. Multiple connection does have restrictions they do not connect or allow all the open requests from web. IoT will connect many devices and all of them will be generating data frequently and to identify or address all the connected devices we will have to use naming mechanisms to trace them. All the connected devices will generate data and to share generated data, will require bandwidth and reliability on services for smoothly communicate these data. Different application requires different strategies. They have different processing capabilities. So, we should have middleware solutions, short-term versus long-term storage requirements. Processing of data will have clustered, summarized and abstract will reduce chock up of data at device/source level. Instead of sending/transferring all the data to network. These allow device to send only knowledge data. These will improve capability and services of applications. However, some of the applications and scenarios will require higher capability and timely data with efficient communication into processing of application and services like disaster monitoring applications.

C. Representation and Publication

Different data publishers represent data in various ways and provide access to sensory data. Sensory data can be stored for short-term or long-term period. Data is accessible and in use till it is stored. As data size and variation in real-world data increase, publishing and querying of the data becomes more challenging. In recent research, there are several researches on adding tags to enhance machine interoperable description on data.

World Wide Web – has proposed few models [4,5] and frameworks for real-world data publication and representation [6]. A research is ongoing on public-

cation and representation for assorted real-world data to automate annotations, mapping schema models. This will also balance complexity of metadata and balance between expressing data streams.

D. Query and Discovery

The query and discovery of real-wold data are based on location, type and other important aspects. But in high volume data are distributed and non-static. Mentioning data description and querying the same over distributed environment isn't an easy task. However large volume data sets provide multiple location and independency on data over quality to real-world spectacles where finding solutions where the data dependency and location varies in required set of data. In real world we are still legging of matured search engine just like web. Web resolve real-time issues, indexing, querying and discovery. To manage other tasks like publishing, understanding, analysing and sharing streaming of data, innovators are extending technologies.

However, there are many researches and suggestion towards streaming data processing of annotated data [7,8]. By enhancing allow constant query over data, WoT application will be more capable of transforming and integrating physical world data along with its description of domain. These can be implemented as Linked Open Data.

E. Access, Subscription and Integration

Real-world data often require for decision making or research purpose. WoT data are more useful if they are stored as individuals rather than combining with different types of data. These combinations create complexity to understand or querying those data. In case if integration becomes automated then it must have cooperation between real-world objects/data and solution for defining the data. Access to data is continuous process and will require some method to support mobility, energy saving method to access the data.

III. DATA TO KNOWLEDGE

WoT data are large storage and they are accessible continuously. Big Data is also large data, but it has parallelization to execute tasks [9]. Real time access to Machine is also important. WoT resources stay joined in network till gets power but in case of power failure can join another network [11]. Discovery mechanism should obtain data from preferable resource and index it. Quality and type of resources is also a challenge. Inconsistency can be there because of change is device and capability. Feedback method would improve quality-based service. Millions of devices will be connected and keep on generating data. Going further this will add large volume of data and to identify these data would require semantic connectivity for consumers. Large volume of data, connectivity will become more knowledgeable for data extraction. Flexibility in analysis, processing models and machine learning will require to collaborate between methods and solutions. This will provide value chain for data so anyone in world data can be extracted, processed and create actionable information and leads to important decisions. Figure 2 demonstrates different steps that can be envisaged for efficient processing and for making use of WoT data.

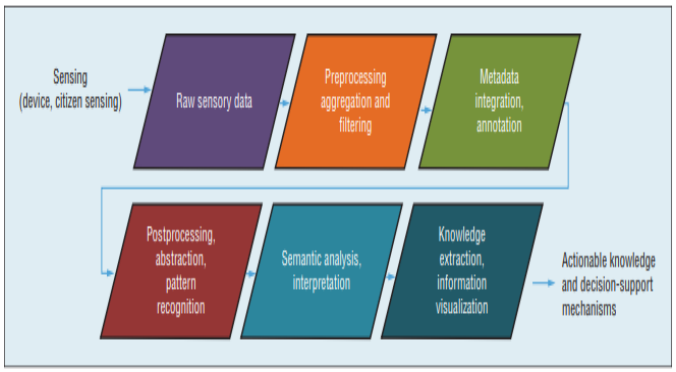

Figure 2: Real-world data process chain.

IV. CONCLUSION

The Research shows how WoT can help to improve real world task by monitoring and interaction with real-world objects. A framework is proposed for

analysis, processing and creating annotated data for alert system by using dynamically provided resources for WoT application. WoT should have integration of network enables devices, M2M and M2H. WoT will produce real-world observations and measurement will introduce more challenges. Due to dynamic resources security and privacy will remain always on top of all issues. Large-volume will have issues on performance of system. Researchers must provide scalable and efficient solutions for real-world data to improve performance and analytics.

Today, WoT is a pouring force to produce, access, and collect data from world. The ability to extract, analyse and interpret this data to create meaningful insights, extract knowledge.

V. REFERENCES

[1] A. Sheth, C. Henson, and S. Sahoo, "Semantic Sensor Web," IEEE Internet Computing, vol. 12, no. 4, 2008, pp. 78–83.

[2] . K. Thirunarayan and A. Sheth, "Semantics-Empowered Approaches to Big Data Processing for Physical-Cyber-Social Applications," Proc. AAAI 2013 Fall Symp. Semantics for Big Data, AAAI, 2013; http://knoesis.org/library/download/aaaiSemanticsAndBigData-TKP-ASPCS.pdf.

[3] A. Sheth, P. Anantharam, and C. Henson, "Physical-Cyber-Social Computing: An Early 21st Century Approach," IEEE Intelligent Systems, vol. 28, no. 1, 2013, pp. 79–82.

[4] M. Compton et al, "The SSN Ontology of the W3C Semantic Sensor Network Incubator Group," J. Web Semantics, vol 17, 2012, pp. 25–32.

[5] L. Lefort et al., Semantic Sensor Network XG Final Report, W3C Incubator Group Report, 2011.

[6] P. Barnaghi et al., "Semantics for the Internet of Things: Early Progress and Back to the Future," Int'l J. Semantic Web and Information Systems, vol. 8, no. 1, 2012, pp. 1–21; doi:10.4018/jswis.2012010101.

[7] A. Bolles, M. Grawunder, and J. Jacobi, "Streaming SPARQL—Extending SPARQL to Process Data Streams," The Semantic Web: Research and Applications, LNCS 5021, Springer, 2008, pp. 448–462.

[8] D. Anicic et al., "EP-SPARQL: A Unified Language for Event Processing and Stream Reasoning," Proc. World Wide Web Conf., ACM, 2011, pp. 635–644.

[9] T. Kraska, "Finding the Needle in the Big Data Systems Haystack," IEEE Internet Computing, vol. 17, no.1, 2013, pp. 84–86.

[10] H.G. Miller, P. Mork, "From Data to Decisions: A Value Chain for Big Data," IT Professional, vol. 15, no. 1, 2013, pp. 57–59.