

From Sensors to Real-time Analytics

Carolina Fortuna, Marko Grobelnik

Jozef Stefan Institute, Ljubljana, Slovenia
E-mail: carolina.fortuna@ijs.si

Abstract. In the paper we present a generic pipeline developed at the Jozef Stefan Institute for the analysis of sensor data. The pipeline consists of various components such as sensors (measuring diverse quantities), sensor boards (installed in smart-city scenarios), communication lines delivering the sensor data to the data center, data servers (to enrich and store the data), real-time analytics (to aggregate and predict the selected signals), and visualization (to deliver information to the user). The pipeline is developed within a series of EU-funded FP7 projects and installed in two smaller towns in Slovenia for testing and development purposes.

Keywords: sensor, VESNA, smart city, real-time analysis, stream mining, complex event detection.

1 INTRODUCTION

One way of looking at the Web/Internet of Things – the way we look at it in our lab – is to see “things” as organs which detect the stimuli [1] [2]. These are then sent via wireless or wired technology, typically on an IP/HTTP network to processing and storage engines. These engines then crunch the received information and generate knowledge. Sometimes they can also trigger an action, such as sending a tweet. This is somehow similar to how we, humans, function: we have five senses which are felt by corresponding organs, then the stimuli are sent to the brain via the nerves, finally the brain processes these stimuli. The result is most often knowledge and perhaps also actions can be triggered: the brain then transmits commands via the nerves to muscles which then trigger moving hands, legs, talking, etc. One distinction is that while in the case of the humans the sensors and processors are spatially close to each other (e. g., nose and brain, ears and brain), in the case of WoT we may be looking at a global distributed system – see Figure 1.

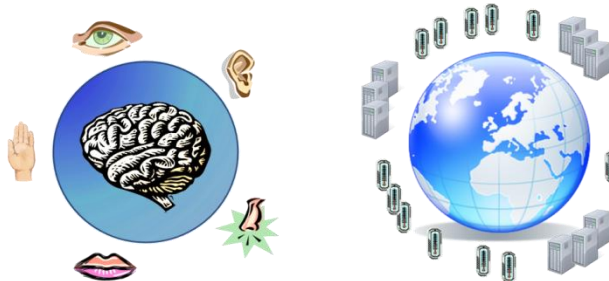


Figure 1: 5 senses analogy of WoT

In this paper, we present a generic pipeline consisting of five main components and its implementation for several smart-town scenarios. We describe the

components of the pipeline and their function in the overall system as well as its application in the identified smart cities. We also discuss the on-going projects which use the components of the pipeline.

This paper is structured as follows. Section 2 details the pipeline and its components. Section 3 describes the smart-town scenarios and the applications based on the pipeline. Section 4 gives an overview of the running projects which use components of this pipeline. Finally, Section 5 summarizes the paper.

2 PIPELINE

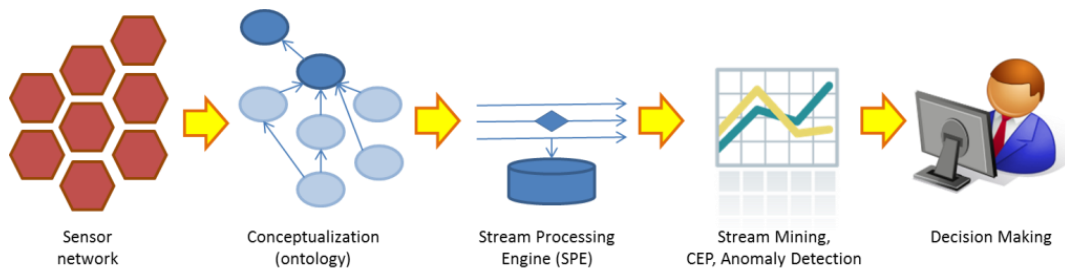
The technological pipeline for the Internet/Web of Things - as we see it - is presented in Figure 2. The pipeline consists of five main components: the network of sensors, the conceptualization of the domain, the stream processing engine, the stream mining and anomaly detection and the consumer [1].

2.1 The sensor network

A subset of the Web/Internet of Things ecosystem is formed by networks of sensors. The data processing part of the pipeline is agnostic to the sensor input, however, the running prototypes are all based on the VESNA sensor nodes¹. The VESNA sensor nodes were designed having in mind a wide array of application areas: energy efficiency, smart transportation, green communications, environmental monitoring, etc.

VESNA is a modular and fully flexible platform for the development of wireless sensor networks developed at the SensorLab @ Jozef Stefan Institute. It is based on a high-performance microcontroller with ARM Cortex-M3 core and radio interface spanning over multiple ISM frequency bands. In terms of modularity, the platform consists of the VESNA core module and a set of special

¹ <http://sensorlab.ijs.si/hardware.html>



Picture 2: Page layout of the document

feature modules (sensor node radio – SNR, sensor node expansion – SNE, sensor node power – SNP) that are used as/if needed.

2.2 The conceptualization of the domain

For small and medium size isolated projects it can be relatively straightforward to know which stream of data measures a given property. Traditional database tables can work well in such situations. However, if we are talking about web scale and aim at interoperability, conceptualization of the Web/Internet of Things domain is needed. The conceptualization plays a key role here by providing a standard way of annotating data streams and meta-data about the things. Standards such as SensorML¹ and Observations and Measurements² coming from the Open Geospatial Forum and the more recent Semantic Sensor Network Ontology (SSN)³ from the W3C are the most relevant when dealing with sensor data in the wide Web/Internet of Things ecosystem.

In our implementation of the pipeline, the SSN ontology is used for describing the sensor platforms and their data⁴. The sensor meta-data is typically collected automatically using the device identification protocol [3]. We took a centralized approach for the meta-data management system, however, we also maintain a smaller demo implementing the distributed approach [4].

2.3 Stream data processing

Smart-city applications are likely to be built on vast amounts of streaming data coming from sensors installed throughout the city's infrastructure as well as from active citizens. In order to make sense of the data, they are likely to need a sensor stream-processing engine able to process high volumes of data with very low latency. This means processing hundreds of thousands of messages per second with a micro/millisecond-range latency. They should be able to take advantage of modern multi-core/multi-processor computer architecture and be able to recover from crash in real-time.

The pipeline uses SenseStream [1] for processing the streams of data.

2.4 Complex event processing

Smart-city applications are also likely to benefit from event detection and processing techniques. A straightforward example would be: sensors in the vicinity of a stadium detect high level of noxes in the air and high level of noise. A complex event processing engine will automatically recognize this as a sports match or concert.

A common characteristic of event processing applications is to continuously receive events from different event sources (e.g. sensors, software modules, blogs, etc.). The central module processing the events, called the CEP engine, detects event patterns from the incoming data streams and outputs the detected or predicted complex events which can be further used by other event consumers, or it can return as an input to the CEP engine.

2.5 Stream mining

Data stream mining is the process of extracting knowledge structures from continuous rapidly-changing streams of data. For the particular case of sensor generated data which often is a time series, stream mining systems extract summaries on which the mining is then performed. Summaries in the form of aggregates or synopsis analysis can be performed on a sliding window over the most recently arrived data. These summaries can be average, standard deviation, percentiles, frequency domain transforms, etc. For instance, Discrete Fourier Transformation (DFT) can be performed on a sliding window to achieve fast subsequence matching. In the frequency space, which in this case has a reduced dimensionality, it is then easier to define a similarity measure between the data from two sliding windows.

SenseStream, the component already used for stream data processing, also handles the stream mining part in the implemented pipeline.

¹ <http://www.opengeospatial.org/standards/sensorml>

² <http://www.opengeospatial.org/standards/om>

³ <http://www.w3.org/2005/Incubator/ssn/ssnx/ssn>

⁴ <http://sensors.ijs.si:2020/>

2.6 Anomaly detection

Anomalies detected in data from the sensor networks can mean one of two things: either sensors are faulty, or they are detecting events that are interesting for further analysis. Sensor data can also be a subject of noise from the environment or can have missing values in the data collection process. In the case of streaming data, anomaly detection techniques are required to perform an online approach, where the techniques need to be lightweight to permit fast processing. Due to the nature of sensor data they also need to distinguish between interesting anomalies and noise/missing values. The key to anomaly detection and successful stream mining is proper feature engineering. Anomalies are detectable if data instances are represented in an informative feature space. Often, it is hard to precisely define what an anomaly is as there are many possible types and there is no universal technique that would detect all. SenseStream performs anomaly detection in the implemented pipeline.

3 TEST DEPLOYMENTS

The pipeline is implemented using several sensor deployments across Slovenia. The data coming from these sensors is then annotated, enriched, processed, mined and fed to several applications. Figure 3 presents a screenshot of Videk [1] – the mash-up application that gives an overview of the deployments. It can be seen that there are currently over 70 nodes installed: 52 VESNAs installed in Logatec and 15 in the municipality of Miren-Kostanjevica. The number of VESNAs installed in Ljubljana changes frequently mostly due to the fact that we have experimental set-ups at the Jožef Stefan Institute campus which are being continuously reconfigured. Currently, there are approximately 16 VESNAs set-ups running in Ljubljana (Videk only shows two because it is not updated with test VESNAs). We also had four VESNAs in Žužemberk and one in Reber.

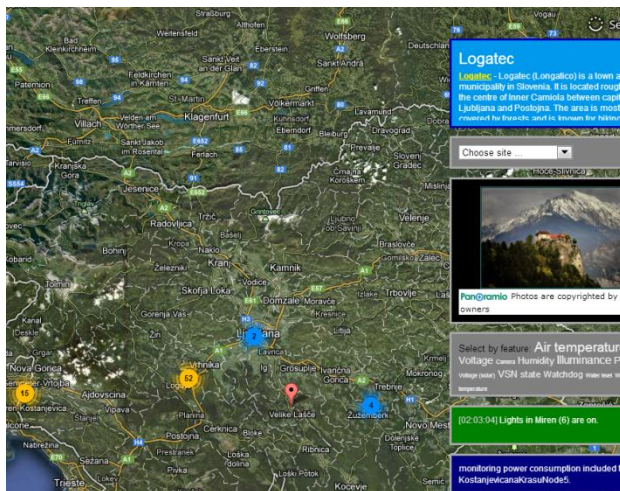


Figure 3: Videk showing VESNA deployments.

3.1 LOG-a-TEC

The LOG-a-TEC wireless sensor network testbed is situated in the town of Logatec, ~30 km from Slovenia's capital Ljubljana. The longer term plan is to extend the LOG-a-TEC testbed with further advanced sensing functionalities and to create new services and applications using ICT for more efficient energy delivery and consumption, environmental and air quality monitoring, traffic monitoring, waste collection and communication systems while preserving or improving the environmental conditions and quality of life.

The first step in this direction was taken by installing 52 VESNA sensor nodes which are capable of sensing the radiofrequency signals present in ISM and UHF bands. With the technology currently developed, we are able to plot radio environmental maps and design next generation green communication systems based on cognitive radio and cognitive networking principles.

3.2 Miren-Kostanjevica

The Miren-Kostanjevica testbed is situated in the municipality of Miren-Kostanjevica, ~100 km from Slovenia's capital Ljubljana. Miren-Kostanjevica will soon benefit from an intelligent energy management system based on, real-time analytics and real-time forecasting services.

The first step in this direction has been taken by setting up an initial environmental monitoring pipeline with focus on studying possibilities for energy saving in the public lighting segment⁵ without affecting the quality of life of the inhabitants.

3.3 Ljubljana

The Ljubljana testbeds can be grouped in two categories. The first category focuses on green communication networks which use green solutions for powering the telecommunication infrastructure. The second category consists of all the VESNA based prototype testbeds which we set up at JSI's campus for testing before field deployment. As a result, this testbed features, among others, an electromagnetic spectrum sensing setup and an environmental monitoring setup.

3.4 Others

Two of our small size experimental setups are located in Žužemberk and Reber. The first set-up focuses on efficient farming by sensing the living conditions in cowsheds. The second setup focuses on intelligent tourism, particularly providing people an information systems which helps them decide when planning sports fishing trips.

⁵ <http://sensors.ijs.si/sl/electricity-saving.html>

4 PROJECTS

We are developing various components of the pipeline in several projects.

4.1 CREW

The main target of FP7-CREW⁶ is to establish an open federated test platform to facilitate experimentally-driven research on advanced spectrum sensing, cognitive radio and cognitive networking strategies for horizontal and vertical spectrum sharing in licensed and unlicensed bands.

The CREW platform incorporates five individual wireless testbeds incorporating diverse wireless technologies (heterogeneous ISM, heterogeneous licensed, cellular, wireless sensor) augmented with State-of-the-Art cognitive sensing platforms. One of these testbeds, the only one in real outdoor environment, is hosted in Logatec and Ljubljana.

4.2 OPCOMM Competence Centre

The OPCOMM Competence Centre⁷ is designing an open communication platform for the development of new cutting-edge services and applications for the Future Internet. Special attention is given to the applicability of services, quality of user experience, the applicable value of data and content, and the interaction with the "material world", i.e. with various devices, objects and processes. This requires efficient interaction between the smart user terminals, appliances and objects, contextually dependent services, and the communication network. The programme encompasses research, design, and prototype development with the final demonstration of new solutions.

The network of sensors and the enrichment components have been significantly improved under the OPCOMM competence center. Security and other connectivity aspects also benefit from this project.

4.3 PlanetData

The PlanetData⁸ project pursues three objectives that altogether enable the creation of a durable community of academic and industrial partners. This community will be supported in conducting research in the large-scale data management area through the provision of data sets and access to tailored data management technology.

From the research point of view, the focus of the project is on large-scale data management. SensorLab provides raw and annotated sensor data and services based on them.

4.4 Envision

The ENVISION⁹ project provides an ENVIRONMENTAL Services Infrastructure with ONTOLOGIES that aims to support non ICT-skilled users in the process of semantic discovery and adaptive chaining and composition of environmental services.

The stream processing and mining components of SenseStream are advancing under this project.

4.5 CITI-SENSE

The CITI-SENSE¹⁰ project will develop "citizens' observatories" to empower citizens to contribute to and participate in environmental governance, to enable them to support and influence community and societal priorities and associated decision making. CITI-SENSE will develop, test, demonstrate and validate a community-based environmental monitoring and information system using innovative and novel Earth Observation applications.

To achieve this, the project will: (i) raise environmental awareness in citizens, (ii) raise user participation in societal environmental decisions and (iii) provide feedback on the impact that citizens had in decisions. It will address effective participation by citizens in environmental stewardship, based on broad stakeholder and user involvement in support of both community and policy priorities. The project aims to learn from citizen experience and perception and enable citizenship co-participation in community decision making and co-operative planning.

The network of sensors is expected to be improved and extended in the direction of environmental and air quality monitoring in public spaces under this project.

4.6 NRG4CAST

The NRG4Cast project is aimed at developing and providing real-time management, real-time analytics and real-time forecasting services for energy distribution networks in urban/rural communities. The four data domains to be analysed are: (1) network topology and devices, (2) energy demand and consumption, (3) environmental data and (4) energy prices.

Besides predictions, the services that will be integrated in the pipeline and final decision support system will include also network monitoring, anomaly detection, route cause analysis, trend detection, planning and optimisation. They will be using advanced knowledge technologies in particular machine learning, data and text mining, stream mining, link analysis, information extraction, knowledge formalisation and reasoning.

All the components of the pipeline are planned to be improved and extended in this project.

⁶ <http://www.crew-project.eu/>

⁷ <http://www.opcomm.eu/en/community/competence-centre>

⁸ <http://www.planet-data.eu/>

⁹ <http://www.envision-project.eu/>

¹⁰ <http://citi-sense.nilu.no/>

5 SUMMARY

In this paper we presented a pipeline enabling the instantiation of the smart-city concept. We described the components of the pipeline, showed example smart-city projects where the pipeline is used and projects in which different components of the pipeline are extended and used.

ACKNOWLEDGMENTS

The authors would like to acknowledge the town of Logatec and the Municipality of Miren-Kostanjevica. This work was partially supported by the Slovenian Research Agency through the P2-0016 programme and J2-4197 project, the KC OPCOMM competence center, and the ICT Programme of the EC under PASCAL2 (ICT-NoE-216886), ENVISION (ICT-2009-249120) and PlanetData (ICT-NoE-257641).

REFERENCES

- [1] C. Fortuna, B. Fortuna, K. Kenda, M. Vucnik, A. Moraru, D. Mladenic: Towards Building a Global Oracle: a Physical Mashup Using Artificial Intelligence Technology , Third International Workshop on the Web of Things (co-located with IEEE Pervasive), June 2012.
- [2] Carolina Fortuna, Marko Grobelnik: The Web of Things, World Wide Web Conference (WWW), Lyon, France, April 2012.
- [3] Matevž Vučnik, Zoltan Padrah, Carolina Fortuna, Mihael Mohorčič: Development of Discovery and Identification Protocol for Sensor Networks, 4th Jožef Stefan International Postgraduate School Students Conference, May 2012.
- [4] C. Fortuna, A. Moraru, P. Oniga, Z. Padrah, M. Mohorcic: Metadata Management for the Web of Things: a Practical Perspective , Third International Workshop on the Web of Things (co-located with IEEE Pervasive), June 2012.
- [5] Klemen Kenda, Carolina Fortuna, Alexandra Moraru, Dunja Mladnic, Blaž Fortuna, Marko Grobelnik (2013) Mashups for The Web of Things. in: Brigitte Endres-Niggemeyer (ed.) Semantic Mashups. Springer (forthcoming).
- [6] Klemen Kenda, Carolina Fortuna, Blaž Fortuna, Marko Grobelnik: Videk - A Mash-up for Enviromental Intelligence, ESWC 2011, May 2011.
- [7] Zoltan Padrah, Tomaž Šolc, Mihael Mohorčič: VESNA based platform for spectrum sensing in ISM bands, 4th Jožef Stefan International Postgraduate School Students Conference, May 2012.
- [8] Matevž Vučnik, Zoltan Padrah, Carolina Fortuna, Mihael Mohorčič: Development of Discovery and Identification Protocol for Sensor Networks, 4th Jožef Stefan International Postgraduate School Students Conference, May 2012.

Carolina Fortuna is a senior research assistant and a PhD student working at the Department of Communication Systems and SensorLab. Her research is interdisciplinary focusing on semantic technologies with applications in modelling of communication and sensor systems and on combining semantic technologies, statistical learning and networks for analyzing large datasets. She has actively participated in FP6 and FP7 projects (FP7 CREW-IP, FP7 ACTIVE-IP, FP7-REGPOT-AgroSense, FP7 PlanetData-NoE, FP6-IST-CAPANINA) and gained industry experience by interning with Bloomberg LP and Siemens PSE.

Marko Grobelnik is an expert in the areas of analysis and knowledge discovery in large complex databases. In particular, the areas of expertise comprise: Data Mining, Text Mining, Semantic Technologies, Network Analysis, and Complex Data Visualization. Marko collaborates with major European and US academic institutions and consults industries such as British Telecom, Microsoft Research, Nature, New York Times, Bloomberg, and Accenture. Marko is author of several books in the area of machine learning, data mining and semantic technologies and authors of many scientific papers. He is also W3C AC representative for IIS, CEO of the company Quintelligence and founder of the Cycorp Europe company.