Semantic web technologies as enablers for truly connected mobility within smart cities

Citation for published version (APA):

Document license:
CC BY-NC-ND

DOI:
10.1016/j.procs.2019.04.008

Document status and date:
Published: 21/05/2019

Document Version:
Publisher’s PDF, also known as Version of Record (includes final page, issue and volume numbers)

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher’s website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
• You may freely distribute the URL identifying the publication in the public portal.

If the publication is distributed under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license above, please follow below link for the End User Agreement:
www.tue.nl/taverne

Take down policy
If you believe that this document breaches copyright please contact us at:
openaccess@tue.nl

providing details and we will investigate your claim.
Semantic web technologies as enablers for truly connected mobility within smart cities

Miloš Viktorović a,*, Dujuan Yang a, Bauke de Vries a, Nico Baken a,b

Abstract

Most car manufacturers predict that in the first half of the next decade there will be fully autonomous vehicles on our roads. Such vehicles would have to communicate in order to mitigate problems caused by single-viewpoint approach. So there are a lot of researches and developments when it comes to communication layer of V2X (Vehicle-to-Everything), but there is still a lot to be done when it comes to data layer of this communication. This is why we propose using Semantic Web Technologies (SWT) to fill in gaps within data layer of V2X communication. By using SWT (Semantic Web Technologies) and Linked data, we plan to interconnect various data sources, in order to provide homogeneous way for connected autonomous vehicles (CAV) to access relevant information. Such information is currently contained in three distinctive type of sources. These are: Geo-stationary Static data sources (Maps, City models), Geostationary Dynamic data sources (IoT devices) and Non-geostationary Dynamic sources (Vehicles). Using SWT, our goal is to develop ontology(s), in such a way that in-vehicle algorithms can extract and process information about environment they are in, while taking into account available network bandwidth.

© 2019 The Authors. Published by Elsevier B.V.

This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Peer-review under responsibility of the Conference Program Chairs.

Keywords: Linked data; Authonomous vehicles; Semantic web; Mobility; Smart Cities; Ontologies;

* Corresponding author. Tel.: +31-(0)62-134-3331;
E-mail address: m.viktorovic@tue.nl
1. Introduction

Accurate understanding of the surrounding environment is crucial for autonomous vehicles (AV). In order to operate in the safest possible way, AVs have to understand and process data from sensors, which by default have certain limitations, due to a single-point view perception. Such challenges arise due to physical obstructions of sensors, together with technical limitations of sensors and processing systems [1] (limited range, uncertainty in classification, bad lightning, etc.). With such a limited information, vehicles cannot perceive the world in the proper way, and therefore their actions will be based on this distorted perception. In order to mitigate this problem, vehicles have to communicate with other vehicles and with traffic infrastructure around them in order to get the best possible information about the environment they are in.

That is why in this paper, we propose utilizing Semantic Web technologies (SWT), to provide a homogeneous machine-readable data structure, which supports CAVs (Connected AVs) operation within urban environments.

In order to demonstrate our proposal, we have structured the paper as follows: Section 2 will provide a short overview of Semantic Web concepts, pointing out examples, where these concepts are used and proven to bring benefits. Next section (Section 3) will then focus on ontology generation, mapping and evolving, as these are crucial steps in the development of linked data structures. Section 4 will point out some challenges of single-viewpoint approach in AVs, and general need for knowledge sharing in mobility. In Section 5 we will present our proposal, on how we plan to use SWT, to support information sharing for CAVs. Also we will briefly describe how we intend to conduct the proposed research.

2. Semantic web and linked data

The term linked data is used for describing concept of data model, based on semantic relationships between stored representation and real world [2]. This term is closely related to semantic web, which is based on semantically linked data. As it is stated by W3C (The World Wide Web) Consortium: “Linked Data lies at the heart of what Semantic Web is all about: large scale integration of, and reasoning on, data on the Web.”[3]

Although created for the purpose of building machine-readable web of data, semantic web technologies found their way in many applications, where they have proved as invaluable for bridging the problem of heterogeneous data sources. Examples of these are linked data applications developed for GIS (Geographic Information Systems) [4] and BIM (Building Information Management) [5], but also for many other fields (medicine, agriculture, etc.). Currently the best example is the project Km4City [6], in which researchers have developed a linked data ontologies and connected applications to cover main domains within smart cities.

Linked data is stored as a set of triplets (mostly), which are representations of binary relationships. These triplets are structured in the form of subject-predicate-object. In addition, relationships are defined through other kind of data structure, to which we refer as ontology. Main technologies related to mentioned data structures are:

- RDF (Resource Description Framework) – Framework for representing semantic data [7], developed by W3C
- SPARQL – RDF query language, with many domain specific implementations like GeoSPARQL[8], C-SPARQL[9], etc.
- OWL (Web Ontology Language) – Language developed for knowledge representation in form of ontologies[10]

Although linked data is not new, and it is attracting growing interest in scientific circles, there are still parts where technology has to mature. In the field of data streams and continuous querying, semantic web is still lacking standardization in tools and methodologies, but ongoing developments look promising [11]–[13].

3. Ontologies

As mentioned in the previous section, ontologies represent a specific data structure, which has a goal of defining relationships between object within linked data model. This shared understanding of specific domain and context is
represented as set of classes, relations, functions and axioms [14]. The term itself comes from philosophy, where it represents the concept of Existence.

Context-awareness is very important when creating semantic data models and ontologies, although it is often taken for granted. Based on definition of context given by Dey: “Context is any information that can be used to characterize the situation of an entity” (5), it becomes obvious that the semantic interlinking of data, cannot be done without properly defining context first. Autonomous vehicles as complex environments, require information from multiple sources, which have overlapping contexts. Such environments can be classified as shared-context environments, as defined by Zimmermann et al. [15]. Therefore shared-context approach will have to be reflected in the development of ontologies, as they will interconnect data sources from multi-domain systems.

In the lifecycle of ontology engineering, we can identify three phases: Generation, Mapping and Evolution phase. First two, are often intertwined, while the phase of evolution usually comes as a final step in the lifecycle. Multiple projects have tackled the topic of ontology development, with different goals, and some of the most mentioned in literature [16], [17] are: InfoSleuth (MCC)[18], SKC (Stanford)[19] and Ontology Learning (AIFB)[20]. All of these research groups developed their own methodologies for ontology engineering, which have been reviewed by Dong et al. in two papers, with a conclusion, that still there are no significant breakthroughs in semi-automatic and automatic ontology generation. In 2013, another group of scientist also compared different methodologies, for ontology engineering, and came to the conclusion that most methods in literature lack sufficient details for practical applications[21]. In the same review, Iqbal et al.[21] proposed the use of METHONTOLOGY[22] method for ontology generation, as the most prominent from all analyzed in the review.

4. Need for knowledge management in mobility

4.1. Levels of Driving Automation in Automated driving systems

SAE International standardization, defines 6 levels of autonomy of vehicle driving systems[23]. In here, we will only define two highest ones:

- Level 4 (High Automation) – “the driving mode-specific performance by an automated driving system of all aspects of the dynamic driving task, even if a human driver does not respond appropriately to a request to intervene”[23]
- Level 5 (Full Automation) – “the full-time performance by an automated driving system of all aspects of the dynamic driving task under all roadway and environmental conditions that can be managed by a human driver.”[23]

4.2. In-vehicle intelligence

If we look at the predictions of automobile industry, when it comes to AVs (Autonomous Vehicles), most state that L4 (Level 4 – High Automation) AV tech will be widely available by year 2020[24]. From the other hand, L5 (Level 5 – Full Autonomy) vehicles might take a bit more time and therefore predictions are not as consistent as for L4 AVs. This de-facto means that we will have a huge inflow of data from these connected vehicles. Intel predicted, back in 2016, that CAV in use will stream around 4000GB of data every day[25]. At the moment research is focusing a lot on communication and network layer of V2V (Vehicle-to-Vehicle) and V2X (Vehicle-to-Everything) information sharing. Although this is necessary, we should not forget about data layer in this system[26]. And this data layer in ITS (Intelligent Traffic System) is what still needs further development.

By doing a literature review, we have identified, that to the best of our knowledge, there is a lack of research on how to integrate all data sources available within the urban environments, in such a way, that this data can be efficiently used by CAVs. Closely related research is the one developed by TTI (Toyota Technological Institute), which contains ontologies for describing maps, Vehicle Control and vehicles themselves[27].

In making vehicle truly autonomous, we can distinguish two groups of approaches[28]:

- Mediated perception approaches – multicomponent system, which takes all sensory inputs and with a help of AI, translates this complex information into driving outputs
• Behavior reflex approaches – Using machine learning algorithms to map direct connections between sensory inputs and driving outputs

As vehicle is a complex system, we cannot talk about one specific algorithm or approach. For every purpose we have different ways of processing information, ranging from more traditional like Model Predictive Control to Deep Learning algorithms, that are attracting a lot of attention recently[26]. Therefore all of this intelligence combined, we can simply refer to as in-vehicle AI.

In order to feed-in all of these systems with information, we rely mostly on sensors in the car itself (single-viewpoint system). This is of course logical step, as these are mission-critical systems and therefore for basic functions AVs have to be self-sustained. Unfortunately, this approach has some drawbacks[1]:

• Sensory limitations – system can only “see” what is unobstructed within the range of sensor

• AI system limitations – Decision making systems rely on pre-programmed knowledge only to decipher sensory inputs, which can lead to misclassification, uncertain detection, etc. in single-viewpoint systems.

In order to mitigate these drawbacks, system for knowledge management and information sharing needs to be established. Data sources should not be limited to in-vehicle data only, but from all possible sources related to traffic. Such sources could range from actors, like statistic and planning bureaus of municipalities and states, to IoT sensors and actuators, used to record or interact with traffic, together with connected vehicles themselves.

In order to make this information useful for AVs, established information sharing would have to be standardized in such a way, that:

• there is a homogenous way of accessing data

• data is coded and formatted in a certain way with a intention to be consumed by machines

In addition, attention has to be paid to network usage. As already mentioned, CAVs will stress bandwidth, therefore every piece of data extra, should be presented in a way that in addition to being useful to AV, should also not pose bigger than necessary stress for the network.

5. SWT application for CAVs

As CAVs can obviously benefit from extended knowledge about their surroundings, we believe that Semantic Web technologies (SWT) can be utilized to efficiently enable data sharing from V2X communication. Main reasons for proposing SWT for this kind of application, although the technology was not created with this in mind, are that SWT can provide:

• Machine-readable data structures

• Ability to interconnect heterogeneous data sources into a homogeneous structure of shared knowledge.

In addition, we have seen that SWT have been utilized in applications for GIS, BIM and IoT (Internet of Things), where standards are under development by OGC (Open Geospatial Consortium). In addition to this, project Km4City has demonstrated that SWT can be beneficial in managing data for Smart City domains [29]. We also believe that currently available governmental and other open data, sensory and static in nature, should be included as an integral part of overall ITS. And as such, we believe that Km4City is giving us a solid ground for our proposal, as Semantic Web holds huge promises for revolutionizing the way we access data.

For this reason, we believe that we can utilize the power of Semantic web to support Self-driving vehicles. The scope of our research would be to develop knowledge sharing capabilities, using SWT, which would focus on autonomous-vehicle, within urban environments, in such a way that we will:

• Support processing of the shared information, by AI (Artificial Intelligence), in the most optimal way, using AV hardware.

• Utilize the network bandwidth in the most efficient way, while satisfying the first condition for sharing of the needed information

In order to achieve these goals, we plan to answer main research question: How can SWT be used to structure data, so that information retrieval operations, performed by vehicle’s AI, would be most optimal from the point of in-vehicle resources and network bandwidth usage? Also we would have to define in which way it should be codified, so that we can achieve two goals stated above.

Before going into ontology development, we should identify sources of information, but what is more important, research has to be done to discover and define, which are the information, that influence the control of AVs, and
where can this information come from. At this stage, for Netherlands we have identified a couple of data sources, from governmental and municipality bodies like: Kadaster[30], NDW (National Data Warehouse for Traffic Information)[31], Talking Traffic Partnership[32], Eindhoven Open Data[33], together with data from other projects, like KITTI[34], and Km4City[6]. Only after concisely answering these questions, we can focus, on our main research goal, which is the development of ontologies, to support CAVs.

At this stage, based on the knowledge collected, we propose modeling and linking of data from three types of sources: static traffic infrastructure, traffic related sensors and actuators and connected vehicles. In order to achieve this, we would develop necessary ontologies in 3 phases:

- Phase 1: Development of ontology to accommodate static part of infrastructure
- Phase 2: Development (extension) of ontology to accommodate dynamic (geo-stationary) part of infrastructure
- Phase 3: Extension of developed ontology(s) to accommodate vehicles’ “floating” (dynamic, non-geostationary) data.

In the process of ontology(s) generation, we would focus on already developed ones, primarily from Km4City[6] and PDOK Labs projects[35], and their mapping, with necessary evolving to support AVs in the best possible way. In addition, we would focus on already developed standards for ITS systems and information sharing, like DATEX II, where we would base our ontology so that straight-forward relations between XML schema of standard and newly developed ontologies could be established. Also we would follow standards for GIS semantic representation of objects and Semantic Sensor web guidelines[4], in order to properly incorporate static and dynamic parts of infrastructure into linked data model.

With all previous in mind, our goal is to have standardized ontology for traffic knowledge sharing, supported by necessary technologies, at the end of the research project. In addition we plan to devise multiple use-cases that would be used for validation of proposed ontologies.

6. Conclusion and further research

The inflow of CAV into our cities, will definitely start a new era, not only in the domain of mobility, but in many others, since it will create a completely new perception of what vehicle is. Huge increase in the amount of data available, will push us to find new ways how to structure and process this data efficiently. We also believe that CAVs will not only be producers, but also consumers of huge amounts, of information, therefore we have chosen to address them.

In order to cope with challenges in the domain of big data, for vehicle usage, we are proposing a development of a system, based on SWT. Such system will aim to homogenize and standardize a way vehicles handle data. We also have a goal of interconnecting all mobility relevant data sources, using the concept of linked data, in such a way, that information can be efficiently extracted, bandwidth wise and efficiently processed by the AI inside the vehicle.

In order to achieve our goals, we plan to rely on already existing developments and standards, so that our proposed system would remain compatible with systems already in place.

References

[Accessed: 30-May-2018].


