Making smart knowledge from not-so-smart data

Nicolas Seydoux, Khalil Drira, Guillaume Garzone, Nathalie Hernandez, Thierry Monteil

Making smart knowledge from not-so-smart data

Nicolas Seydoux*[1, 2, 3], Khalil Drira†[1, 2], Guillaume Garzone‡[1, 2], Nathalie Hernandez§[3], and Thierry Monteil¶[1, 2]

[1]LAAS-CNRS, 7 av. du Colonel Roche, F-31400 Toulouse, France
[2]Univ de Toulouse, INSA, LAAS, F-31400, Toulouse, France
[3]IRIT, Maison de la Recherche, Univ. Toulouse Jean Jaurès, 5 allées Antonio Machado, F-31000 Toulouse

1 Motivational use case

Smart cities are by essence cyber-physical systems of systems: their emergence is fueled by the joint deployment of devices networks and data-driven services. Very diverse area of urban living are of concern in smart cities: both public and personal transportation systems, resources management, including water and electricity, environment monitoring for weather and pollution, safety of the citizens crisis management in emergency situations, entertainment with tourism travelling and cultural information... All these areas have heterogeneous stakeholders and end users, they are based on unrelated infrastructures, and their data sources are a permanently evolving mix of pre-existing datasets, devices, and human input remotely through social networks and locally through physical interaction.

Let us consider the daily life of Julia, a smart city citizen. When Julia gets on the bus, she uses her smartphone to pay for the bus fare, and the smartphone communicates locally with the surrounding on-board information sources or remotely with cloud applications, in order to discover points of interest that are on the bus route, and especially near the bus stops where she is most likely to get off. As it happens, there is a new exposition of a painter she likes, at a local museum. There is also a convenience store that sells her usual brand dog food, and she is almost out of it. When she approaches her home, she notices that a street lamp is dysfunctional, and notifies on a city smartphone application.

*nicolas.seydoux@laas.fr
†khalil.drira@laas.fr
‡guillaume.garzone@laas.fr
§nathalie.hernandez@irit.fr
¶thierry.monteil@laas.fr

When she gets home, the heating system raised the temperature after it has been kept at a lower level while she was out.

2 Deconstructing the use case

Underlying this use case are multiple challenges and design choices. First, one can notice the user-centric approach, with a high adaptation of the service to the customer profile. Data is integrated from very diverse sources: a public repository for the museum, a company one for the convenience store, and a private one for the information that Julia was out of dog food. This part also raises privacy issues, because not all of her application should have access to sensitive data such as her position or her habits. Julia also becomes a human sensor when she provides feedback about the streetlamp.

To sum it, this use case is grounded into highly interoperable, yet very flexible, machine-understandable data. It supposes privacy policies explicit description, and actionable data that can be used as an input to many application in diverse domains. The semantic web is one way to accomplish this vision.

3 Building smart data

The semantic web proposes highly expressive, dereferencable data descriptions based on vocabularies called ontologies. This expressivity allows to give comprehensive descriptions of devices, either embedded in the device itself or in a description repository (called a knowledge base). Semantic web technologies also allow the production of rich knowledge, capturing its own collection context to make it more reusable.

The high level of formalism of semantic models enables knowledge to be leveraged in reasoning processes, where new knowledge is inferred based on existing knowledge and rules. These rules can be either generic, pre-existing rules, or application specific. The formalism of semantic knowledge also allows for inconsistency detection, which is a real challenge in the domain of smart city where a piece of data will be used in many various domains.

4 Open challenges

The high expressivity and formalism of the semantic web technologies and principles come at a cost, and their adoption as solutions in the constrained domain of the IoT still faces challenges. Richer data formats are more resource-intensive for processing and memory, while the IoT is characterized by the high distribution and high constraints on the nodes. Adaptation of the data is therefore required depending on the node that is processing it. With 50 billion connected nodes predicted by 2020, and the obvious increase in term of produced data volume it entails, scalability is a core concern for semantic solutions to IoT issues.