ECSEL Joint Undertaking
Electronic Components and Systems for European Leadership

MANTIS

Cyber Physical System based Proactive Collaborative Maintenance

D1.2 Consolidated State-of-the-Art of Sensor-based Proactive Maintenance Appendix 11:

Data unification and semantic reasoning for integration of data from heterogeneous sources

Work Package	WP1 - Service platform architecture requirement definition. Scenarios and use cases descriptions
Version	1.0
Contractual Date of Delivery	30/04/2016
Actual Date of Delivery	03/06/2016
Dissemination Level	Public
Responsible	Erkki Jantunen
Contributors	Tom Tourw□ (SIR), Elena Tsiporkova (SIR), Mathias Verbeke (SIR), Tom Ruette (SIR), P□l Varga - AITIA, Csaba Hegedűs – AITIA, István Moldov□n - BME

The MANTIS consortium consists of:

Num.	Short Name	Legal Name	Role	Country
1	MGEP	Mondragon Goi Eskola Politeknikoa J.M.A. S.Coop.	CO	ES
2	MONDRAGON	Mondragon Corporacion Cooperativa S.Coop.	BEN	ES
3	IKERLAN	Ikerlan S.Coop.	BEN	ES
4	TEKNIKER	Fundacion Tekniker	BEN	ES
5	FARR	Fagor Arrasate S.Coop.	BEN	ES
5.1	KONIKER	Koniker S.Coop.	TP	ES
6	GOIZPER	Goizper S.Coop.	BEN	ES
7	ACCIONA	Acciona Infraestructuras S.A.	BEN	ES
8	MSI	Mondragon Sistemas De Informacion S.Coop.	BEN	ES
9	VTT	Teknologian Tutkimuskeskus VTT Oy	BEN	FI
10	LUAS	Lapin Ammattikorkeakoulu Oy	BEN	FI
11	NOME	Nome Oy	BEN	FI
12	FORTUM	Fortum Power And Heat Oy	BEN	FI
13	SQ	Solteq Oyj	BEN	FI
14	WAPICE	Wapice Oy	BEN	FI
15	AAU	Aalborg Universitet	BEN	DK
16	DANFOSS	Danfoss A/S	BEN	DK
17	UNIV	Universal Foundation A/S	BEN	DK
18	HGE	Hg Electric A/S	BEN	DK
19	VESTAS	Vestas Wind Systems A/S	BEN	DK
20	SIRRIS	Sirris Het Collectief Centrum Van De Technologische Industrie	BEN	BE
21	ILIAS	Ilias Solutions Nv	BEN	BE
22	ATLAS	Atlas Copco Airpower Nv	BEN	BE
23	3E	3e Nv	BEN	BE
24	PCL	Philips Consumer Lifestyle B.V.	BEN	NL
25	PHC	Philips Medical Systems Nederland B.V.	BEN	NL
26	PHILIPS	Philips Electronics Nederland B.V.	BEN	NL
27	S&T	Science and Technology B.V.	BEN	NL
28	TU/E	Technische Universiteit Eindhoven	BEN	NL
29	RUG	Rijksuniversiteit Groningen	BEN	NL
30	UNINOVA	UNINOVA - Instituto de Desenvolvimento de Novas Tecnologias	BEN	PT
31	ISEP	Instituto Superior de Engenharia do Porto	BEN	PT
32	INESC	Instituto de Engenharia de Sistemas e Computadores do Porto	BEN	PT
33	ADIRA	ADIRA - Metal Forming Solutions S.A.	BEN	PT
34	ASTS	Ansaldo STS S.p.A.	BEN	IT
35	CINI	Consorzio Interuniversitario Nazionale per l'Informatica	BEN	IT
36	AIT	Austrial Institute of Technology GmbH	BEN	AT
37	HBM	Hottinger Baldwni Messtechnik GmbH	BEN	AT
38	INNOTEC	Innovative Technology and Science Limited	BEN	UK
39	AITIA	AITIA International Inc.	BEN	HU
40	BME	Budaperst University of Technology and Economics	BEN	HU
41	JSI	Josef Stefan Institute	BEN	SI
42	XLAB	XLAB d.o.o.	BEN	SI
43	FHG	Fraunhofer Institute for Experimental Software Engineering IESE	BEN	DE
44	M2X	M2Xpert GmbH & Co KG	BEN	DE
45	STILL	STILL GMBH	BEN	DE
46	BOSCH	Robert Bosch GMbH	BEN	DE
47	LIEBHERR	Liebherr-Hydraulikbagger GmbH	BEN	DE

Document Revisions & Quality Assurance

Revisions:

Version	Date	By	Overview
0.1	6.8.2015	Mathias Verbeke (SIR)	First draft
0.2	28.09.2015	Istv□n Moldov□n (BME)	Added telecommunication related SOTA
0.3	06.10.2015	Riku Salokangas	Added contractual date of delivery etc.
0.4	30.03.2016	Tom Ruette (SIR)	Related standards
0.5	01.04.2016	Istv□n Moldov□n (BME)	Related standards complemented
1.0	02/06/2016	Mikel Muxika (MGEP)	Format correction Deliverable info update

Abstract

Data unification is concerned with combining data from different types, levels and sources in such a way that they are made compatible and comparable, and thus useful for further processing. This is particularly relevant in the multi-stream context considered in MANTIS. Furthermore, there are also differences in the interpretation of the meaning of data, i.e., semantic heterogeneity. Semantic models deal with this issue by offering of way of formally representing knowledge about a particular domain, thereby defining a controlled vocabulary. Subsequently, using semantic reasoners, the ontological constructs can be employed to infer additional knowledge. This appendix offers an overview of current approaches for data unification, and semantic modelling and reasoning.

Table of Contents

1 Introduction

Semantic web technologies provide possibilities to define □ontologies□ - formal data models which have several benefits like heterogeneous data integration, abstraction mechanisms which come handy in a complex multi-context maintenance scenario. In this document we overview the most important technologies that enable data unification and formal reasoning, and present previous work done in this area.

2 Data unification and semantic reasoning for integration of data from heterogeneous sources

Data unification is concerned with combining data from different types, levels and sources in such a way that they are made compatible and comparable, and thus useful for further processing. This is particularly relevant in the multi-stream context considered in MANTIS, as first of all data can originate from thousands of sensors, distributed over different components and systems from different manufacturers, models and versions, and is thus not easily comparable. Second, data has been collected for different purposes using different standards and methodologies, is not captured in a uniform way (e.g., it can be discrete, e.g., log events, or continuous, e.g., sensor readings), is logged at different time intervals (e.g., every hour) with varying temporal granularity (e.g., over 15, 30 or 60 minute time intervals) and with different scales and formats (e.g., time zones), missing values occur for unknown reasons, and values can originate from diverse sources (e.g., from sensors, models or entered manually).

In the past decades, data interoperability and data integration have been extensively explored research topics, mainly focused on syntactic (e.g., differences in data formats) and structural (e.g., differences in data models) data heterogeneity. A major remaining challenge is semantic heterogeneity, e.g., differences in the interpretation of the meaning of data. A promising approach to semantic integration is the use of a semantic model, which is, in general, a way of formally representing knowledge about a particular domain, thereby defining a controlled vocabulary. Ontologies are the most popular form of semantic models, but other forms exist.

Semantic models are used for various purposes and in many domains. They have been used most successfully for data integration purposes in the life sciences domain, where the need for controlled vocabularies is high, due to the huge amounts of experimental data that are generated and the wide variation in terminology that is used. Two of the most well-known initiatives are the Gene Ontology [1], aimed at standardizing the representation of genes and their products across species and databases, and the Open Biological and Biomedical Ontologies [2], a collection of 60 life science-related ontologies. In recent years, they are also increasingly used in the ubiquitous computing domain for realizing context-aware applications, i.e., applications that take into account and adapt to the (physical, computational and user-specific) context in which they operate. Examples of ontology-based context modeling approaches can be found in [3] and [4]. Only recently are they being used in the context of the topic of this project, for example in the sustainable energy domain. [5] made an initial effort to bring the concepts of the PV domain together in an ontology. An additional effort for constructing a PV ontology can be found in [6]. Not surprisingly, these efforts are often related to advances in the use of Complex Event Processing systems (see below in the state of the art on High-performance stream-based processing) that have to deal instantaneously with the formal variety in the data streams. As an example, [7] argue for the adoption of ontologies for easier pattern specification, and [to] detect inexact patterns. A similar argument is made in [8]. For wind turbines, the situation is less dire than for PV systems. In 2009, [9] proposed an ontology model for wind turbines, specifically for condition monitoring. [10] published on predictive diagnosis on the basis of a fleet-wide ontology. More recently, [11] investigated intelligent fault diagnosis of wind turbines on the basis of an ontology.

Besides the huge collection of existing semantic models, little work has been done on the development of ontologies to support the monitoring of networks of heterogeneous and collaborative systems. In order to build an ontology for multi-stream data analysis in this context, ontology co-creation can be used in order to involve stakeholders, who will use the ontology-based decision support platform, in the process of ontology engineering. Ontology co-creation encourages users to have a sense of control over the ontology, and holds the key to solving the challenge of creating an ontology that is both accurate and useful [12]. While the benefits of ontology co-creation are straightforward, its methodological implementation is not. HCOME [13], a methodological approach to involve domain experts into ontology engineering, is highly demanding in terms of the effort it requires from these experts.

Several tools exist that support both the development of these ontologies and their integration into decision support applications. The most popular ontology editor is Prot□g□ [14], a pluggable editor developed at Stanford University. Jena [15], OWL-API [16], Pellet [17], and EYE [18] are semantic reasoning frameworks that offer APIs for programmatically using, creating, manipulating and serializing ontologies. Using these semantic reasoners, the ontological constructs can be employed to infer additional knowledge such as fleet operation and maintenance decisions. First steps are also made to semi-automatically generate charts from queries over these semantic data [19].

In network management and monitoring several tools and platforms are used to gather measurement results. Each platform uses its own data structures and its own interaction interfaces, therefore integration of the gathered data is difficult.

One possibility for integration is to normalize the measurement data to a predefined XML syntax and define interfaces to access the normalized data. Such solution would be to use the Open Grid Forum Network Measurements Working Group (NMWG) [20].

To accommodate heterogeneous management platforms, semantic technologies can be used to enable abstract treatment. At the same time, it is possible to define the information at different abstraction levels, which allows the definition of specific class of measurements that are derived from generic ones.

The FP7-MOMENT (Monitoring and Measurement in the Next Generation Technologies) project [21] applies the concepts provided by ontologies to address the integration of measurement information from a semantic viewpoint. MOMENT proposal for traffic measurement data and metadata ontologies takes into account previous works (eg. OGF-NMWG [22], W3C Time [23], NASA Units [24]), which were adapted to the mediator requirements. Thus, a core ontology has been built from them and from all databases schemas the mediator connects to. The metadata ontology describes the available repositories and measurements: where they are and what data they contain. It was derived mainly from metadata schemas, such as CAIDA's DatCat [25] or MOME [26] project.

Another semantic-based approach to the dynamic monitoring, analysis and control of heterogeneous telecommunication networks is the Bell Labs□ SNoMAC (Semantic Network Monitoring, Analysis and Control) research [27]. The SNoMAC solution formally describes networked devices based on a layered collection of extensible ontologies grounded in an upper-level ontology called NetCore. Together these ontologies make it possible to hide or expose device details to the extent necessary for the problem in hand. In this way device- and protocol-specific information can be semantically encapsulated at the lowest levels, with higher layers focusing on more generic networked device characteristics. As a result, applications can be written against SNoMAC□s formal semantic API, ignoring lower-level details that may change as standards evolve and new ones are added.

3 Conclusions

Data unification and semantic reasoning for integration of data from heterogeneous sources is definitely a requirement for a diverse scenario like in MANTIS. A formal description which shows a generic interface while hides the low level details needs to be developed, where we can exploit the advantages of semantic reasoning as well.

4 Related standards

- ieee 1636: IEEE Standard for Software Interface for Maintenance Information Collection and Analysis (SIMICA), including 1636.*

- iso 13372: Condition monitoring and diagnostics of machines -- Vocabulary

- bsi 13306: Maintenance terminology

- UNI 10144: Classification of maintenance services

- ETSI TR 101 584: Machine-to-Machine communications (M2M); Study on Semantic support for M2M Data

- ETSI TS 103 264 SmartM2M; Smart Appliances; Reference Ontology and oneM2M Mapping

- ISO 13374-1 Condition monitoring and diagnostics of machines -- Data processing, communication and presentation -- Part 1: General guidelines

- ISO 13374-2 Condition monitoring and diagnostics of machines -- Data processing, communication and presentation -- Part 2: Data processing

- ISO 13374-3 Condition monitoring and diagnostics of machines -- Data processing, communication and presentation -- Part 3: Communication

- ISO 13374-4 Condition monitoring and diagnostics of machine systems -- Data processing, communication and presentation -- Part 4: Presentation

- ISO 15926-1 Industrial automation systems and integration -- Integration of life-cycle data for process plants including oil and gas production facilities -- Part 1: Overview and fundamental principles

- ISO 10303-11 Industrial automation systems and integration -- Product data representation and exchange -- Part 11: Description methods: The EXPRESS language reference manual

References

[1] http://www.geneontology.org/

[2] http://www.obofoundry.org/

[3] Harry Chen, Filip Perich, Timothy W. Finin, Anupam Joshi: SOUPA: Standard Ontology for Ubiquitous and Pervasive Applications. MobiQuitous 2004: 258-267

[4] Xiaohang Wang, Daqing Zhang, Tao Gu, Hung Keng Pung: Ontology Based Context Modeling and Reasoning using OWL. PerCom Workshops 2004: 18-22

[5] Antonino Piazza, Giuseppe Faso: Concentrated Solar Power: Ontologies for Solar Radiation Modeling and Forecasting. Advances onto the Internet of Things: 325-337 (2014)

[6] Fereidoon P. Sioshansi: So what□s so smart about the smart grid? The electricity journal 24(10): 91-99 (2011)

[7] Qunzhi Zhou, Yogesh Simmhan, Viktor K. Prasanna: Towards an inexact semantic complex event processing framework. DEBS 2011: 401-402

[8] Andreas Wagner, Sebastian Speiser, Oliver Raabe, Andreas Harth: Linked Data for a Privacy-aware Smart Grid. GI Jahrestagung (1) 2010: 449-454

[9] Panagiotis Papadopoulos, Liana Cipcigan: Wind turbines□condition monitoring: an ontology model. SUPERGEN 2009: 1-4

[10] Gabriela Medina-Oliva, Alexandre Voisin, Maxime Monnin, Jean-Baptiste Leger: Predictive diagnosis based on a fleet-wide ontology approach. Knowl.-Based Syst. 68: 40-57 (2014)

[11] Anmei Zhou, Yu Dejie, Wenyi Zhang: A research on intelligent fault diagnosis of wind turbines based on ontology and FMECA. Advanced Engineering Informatics (2014)

[12] C. E. Kuziemsky, F. Lau: A four stage approach for ontology-based health information system design. Journal of Artificial Intelligence in Medicine, Vol. 50, Nr. 3, pp. 133-148, 2010

[13] K. Kotis, G. Vouros: Human-centered ontology engineering: the HCOME methodology. Knowledge and Information Systems, Vol. 10, Nr. 1, pp. 109□131, 2006

[14] http://protege.stanford.edu/

[15] http://openjena.org/

[16] http://owlapi.sourceforge.net/

[17] Evren Sirin, Bijan Parsia: Pellet: An OWL DL Reasoner. Description Logics 2004

[18] Jos De Roo: Euler proof mechanism (1999□2014), http://eulersharp.sourceforge.net/

[19] Marcello Leida, Du Xiaofeng, Paul Taylor, Basim Majeed: Toward Automatic Generation of SPARQL Result Set Visualizations - A Use Case in Service Monitoring. ICE-B 2011: 181-186

[20] J. Zurawski, M. Swany, D. Gunter: A Scalable Framework for Representation and Exchange of Network Measurements. In Proc. 2nd International IEEE/Create-Net Conference on Testbeds and Research Infrastructures for the Development of Networks and Communities, Tridentcom 2006, Barcelona, Spain (2006)

[21] Monitoring and Measurement in the Next Generation Technologies (MOMENT). http://www.fp7-moment.eu/

[22] OGF Network Measurement Working group (NMWG): http://nmwg.internet2.edu/

[23] W3C Time Ontology in OWL: http://www.w3.org/TR/owl-time/

[24] Units Ontology: http://sweet.jpl.nasa.gov/ontology/units.owl

[25] Shannon, C, Moore, D., Keys, K., Fomenkov, M., Huffaker, B., Claffy, K.: The Internet Measurement Data Catalog. ACM SIGCOMM Computer Communications Review (CCR) 35(5) (October 2005) 97-100

[26] Gutierrez, P.A.A., Bulanza, A., Dabrowski, M., Kaskina, B., Quittek, J., Schmoll, C, Strohmeier, F., Vidacs, A., Zsolt, K.S.: An Advanced Measurement Meta-Repository. In: Proceedings of 3rd International Workshop on Internet Performance, Simulation, Monitoring and Measurement; IPS-MoMe 2005, Warsaw, Poland (March 2005)

[27] Christopher J. Matheus, Aidan Boran, Dominic Carr, Rem Collier, Barnard Kroon, Olga Murdoch, Gregory M. P. O'Hare, and Michael O'Grady. 2012. Semantic network monitoring and control over heterogeneous network models and protocols. In *Proceedings of the 8th international conference on Active Media Technology* (AMT'12), Runhe Huang, Ali A. Ghorbani, Gabriella Pasi, Takahira Yamaguchi, and Neil Y. Yen (Eds.). Springer-Verlag, Berlin, Heidelberg, 433-444. DOI=10.1007/978-3-642-35236-2˙43 http://dx.doi.org/10.1007/978-3-642-35236-2˙43