

Contents

I Front matter

Abstract	i
Zusammenfassung	iii
Declaration	v
List of Figures	xiii
List of Tables	xv
Code Listings	xvii
Acronyms	xix

II Main matter

1 Introduction	1
1.1 Structure of this thesis	2
1.2 Nomenclature and conventions	3
1.2.1 Figures' informative boxes	3
1.2.2 Prefixes and namespaces	3
1.2.3 Rule language	3
1.2.3.1 Rules in a nutshell	4
1.2.3.2 Example rule	4
1.2.3.3 About rules in this document	5
1.2.4 Text abbreviations	6
1.3 Motivation	6
1.4 Problem context	9
1.5 Goal	11
1.6 Solution requirements	12
1.7 Background	13
1.7.1 Differences between Data, Information and Knowledge	13
1.7.1.1 An illustrative example of DIKW	13
1.7.2 Knowledge Representation	15
1.7.2.1 Syntax	15
1.7.2.2 Semantics	16
1.7.2.3 Pragmatics	16
1.7.2.4 Ontology	16
1.7.2.5 Graph data	16
1.7.3 Data streams	17
1.7.3.1 Stream processing	17

vii

1.7.3.2	Stream reasoning	17
2	Methodology	19
2.1	Design Science Research methodology	19
2.1.1	Types of artifacts	20
2.1.2	Research framework	20
2.1.2.1	Design cycle	21
2.2	Research Questions	22
2.2.1	Main Research Question	22
2.2.2	Knowledge Questions	22
2.2.2.1	Problem-related KQ	22
2.2.2.2	Test-related KQs	23
2.3	Research methods	23
2.4	Assumptions and other aspects	23
3	Problem investigation	25
3.1	Related areas	25
3.2	Semantization of data streams	26
3.2.1	Semantization of individual observations	27
3.2.1.1	Data models at low-level of abstraction	28
3.2.2	Semantization of analytical outcomes	29
3.2.2.1	Data models at high-level of abstraction	30
3.2.3	Other modeling approaches	30
3.2.3.1	AUTOSAR	30
3.2.3.2	NGSI-LD information model	30
3.2.3.3	ASAM standards	30
3.3	Analysis of data streams	30
3.3.1	Statistical summaries	31
3.3.1.1	Data compression using statistics	32
3.3.2	Machine Learning methods on streams	32
3.3.2.1	Deep learning	32
3.3.2.2	A note on datasets for ML	33
3.3.3	Stream mining	33
3.3.4	Stream processing with windowing functions	33
3.3.5	Stream reasoning	34
3.4	Requirements satisfaction of existing approaches	35
3.5	A note on information systems	36
3.5.1	Steaming platforms	36
3.5.2	Cloud services for connected vehicles	38
3.5.3	Graph databases	38
4	Solution approach	39
4.0	Approach overview	39
4.0.1	A zoomed-out view of the solution approach	39
4.0.2	A zoomed-in view of the solution approach	42
4.1	Semantic model	42
4.1.1	VSSo 2.0: A core ontology in sync with VSS	43
4.1.2	XDVP: An ontology eXtension for Dynamic Vehicle Properties	45
4.1.2.1	Scope	45
	Domain & Intended use.	45
	Competency questions.	46
4.1.2.2	Reuse	46

	<i>Reusing VSSo as the basis.</i>	46
	Reusing SSN and SOSA for annotation of individual ob- servations.	47
	Reusing IoT-Stream for time intervals and streams.	47
	Reusing QUDT for units and quantity kinds.	47
4.1.2.3	Terms	48
4.1.2.4	Classes	48
4.1.2.5	Properties	48
4.1.2.6	Constraints	49
4.1.2.7	Instances	49
4.1.2.8	Joining all modeling pieces together	50
	<i>Overall view of entities for static and dynamic data.</i>	50
	<i>Data streams and their features.</i>	50
	<i>Introducing variable kinds.</i>	52
	<i>Data records hold the values of a data stream.</i>	52
	<i>Application produces or consumes a data stream.</i>	52
4.2	Stream analysis	53
4.2.1	Data stream management	54
4.2.2	Analysis of numerical data streams	55
4.2.2.1	Exact mean and standard deviation	55
4.2.2.2	Trend detection with dynamic threshold	56
4.2.2.3	Encoding trends and temporal information	56
	Trend letter and words	57
	Trend codes and their sum	57
	Decision difference and their sum	57
	Ratio of slopes and their sum	57
	Temporal mask	57
4.2.2.4	Symbolic representations	57
4.2.2.5	Feature vectors for complex models	57
4.2.3	Analysis of categorical data streams	58
4.3	Semantic annotation	58
4.3.1	Annotation of numerical data streams	58
4.3.2	Annotation of categorical data streams	59
4.4	Knowledge formalization	59
4.4.1	Extracted concepts from a numerical data stream	59
4.4.2	Extracted concepts from a categorical data stream	59
4.4.3	Examples	59
5	Validation	61
5.1	Validating the VSSo core model	61
5.1.1	Expert opinion on VSSo core	61
5.1.2	Query comparison between VSSo 1 vs. VSSo 2	61
5.1.2.1	ExQ 1: How many seats does this <i>Vehicle</i> have?	62
5.1.2.2	ExQ 2: What is the local current temperature on the driver side?	63
5.1.2.3	ExQ 3: What properties in this vehicle refer to “ <i>speed</i> ”?	64
5.2	Validating the XDVP model	65
5.2.1	Expert opinion on XDVP	65
5.2.2	Competency questions’ satisfaction with SPARQL queries	65
5.2.2.1	Common selection queries	65
	<i>Selecting a specific Vehicle.</i>	65
	<i>Selecting the latest VehicleJourney.</i>	65

	<i>Selecting a specific DataStream of the latest VehicleJourney.</i>	66
5.2.2.2	Answering about background knowledge (i.e., static)	66
	<i>How many driving journeys does a vehicle have in a particular period?</i>	66
	<i>What are the topics of the available streams?</i>	67
	<i>What analytics are applied to a particular stream?</i>	67
	<i>Who are the producers and consumers of a particular stream?</i>	68
	<i>What features are present in a given stream, and what are their quantity type and unit associated?</i>	68
5.2.2.3	Answering about data stream values	68
	<i>What data records of a data stream are available in the latest journey?</i>	68
	<i>How long has the latest seen value of a given categorical feature remained unchanged?</i>	69
	<i>What is the value of a feature (e.g., average) in a given time window?</i>	69
5.3	Analyzing the use of resources	69
5.3.1	Annotations per stream	69
5.3.1.1	Annotations of a categorical stream	70
5.3.1.2	Annotations of a numerical stream	70
5.3.2	Graph data volume comparison	71
5.3.2.1	Graph data sizes from a categorical data stream	72
5.3.2.2	Graph data sizes from a numerical data stream	74
5.4	Building the data graph	75
5.5	Complex prediction tasks	77
6	Use case one: Accessing high-level information inferred from vehicle data streams	79
6.1	Use case overview	79
6.2	Reasoning tasks	80
6.2.1	Vehicle motion	80
6.2.2	Semantic speed	81
6.2.3	Idle mode	82
6.2.4	Road slope	83
7	Use case two: Supporting decision making on the highway	85
7.1	Use case overview	85
7.2	Reasoning tasks	86
7.2.1	Distance-based states	87
7.2.2	Safety conditions' satisfaction	89
7.2.3	Decision recommendation	90
8	Summary	91
8.1	Answers to Research Questions	91
8.1.1	Answer to main Research Question	91
8.1.2	Answer to Knowledge Questions	93
8.1.2.1	Problem-related KQ	93
8.1.2.2	Test-related KQs	94
8.2	Discussion	96
8.2.1	Advantages of using the proposed design	96
8.2.2	Limitations in the proposed design	96
8.3	Future work	97

III	Back matter	
A	RML mapping rules	103
B	Graph data experiment test results	107
	Bibliography	113