

Contents

Part I Overview

1	Introduction	3
1.1	Motivation	5
1.1.1	Adaptability	5
1.1.2	Abstraction	6
1.1.3	Single System Image	6
1.1.4	Complexity of Scientific Applications.....	6
1.1.5	Workflow Modularization	6
1.1.6	Standard Based Graphical Tools	7
1.1.7	Data Flow Optimization	7
1.1.8	Semantics Based Scientific Workflow Composition	7
1.1.9	Automatic Scientific Workflow Composition	8
1.2	Research Goals	8
1.2.1	Abstract Workflow Description Language	8
1.2.2	Workflow Modularization	9
1.2.3	UML-Based Scientific Workflow Modeling.....	9
1.2.4	Data Flow Optimization	10
1.2.5	Semantics-Based Scientific Workflow Composition	10
1.2.6	Automatic Scientific Workflow Composition	10
1.3	Organization	11
1.3.1	Part I: Overview	11
1.3.2	Part II: Programming	11
1.3.3	Part III: Optimization.....	12
1.3.4	Part IV: Synthesis	12
1.3.5	Part V: Related Work	13
1.3.6	Part VI: Conclusions	13
1.3.7	Appendices.....	13

2	Prerequisites	15
2.1	Distributed Systems	15
2.1.1	Programming Architecture	15
2.1.2	Characteristics	17
2.2	Programming for Distributed Systems	17
2.2.1	Programming Issues	17
2.2.2	Programming Models	18
2.3	Scientific Workflow	21
2.3.1	Comparison of Scientific Workflows with Business Workflows	21
2.3.2	Scientific Workflow Basics	23
2.4	Semantics	24
2.4.1	Ontology	24
2.4.2	Web Ontology Language	25
2.5	ASKALON	25
2.5.1	Workflow Composition	26
2.5.2	Resource Management	26
2.5.3	Workflow Scheduling	27
2.5.4	Workflow Execution	27
2.5.5	Performance Analysis and Prediction	27
2.6	Summary	28

Part II Programming

3	Abstract Workflow Description Language	31
3.1	System Overview	31
3.2	Specification	32
3.2.1	Atomic Activity	33
3.2.2	AWDL Control Flow	35
3.2.3	AWDL Data Flow	47
3.2.4	Properties and Constraints	52
3.2.5	The Structure of an AWDL Workflow	53
3.3	AWDL Modeling: A Real-World Scientific Workflow Example	54
3.4	Reification and Execution	57
3.5	Summary	60
4	Workflow Modularization	63
4.1	Introduction	63
4.2	Sub-workflows	63
4.3	Workflow Library	64
4.4	Invocation of (Sub-)workflows	65
4.4.1	Detection of Infinite Sub-workflow Invocation	67
4.4.2	Workflow Invocation	69
4.5	Case Study: WIEN2k	70
4.6	Summary	73

5	UML-Based Scientific Workflow Modeling	75
5.1	Introduction	75
5.2	UML Activity Diagram	75
5.3	Modeling Scientific Workflows with the UML Activity Diagram	76
5.4	Graphical Scientific Workflow Composition Tool	79
5.4.1	Graphical User Interface	80
5.4.2	Model Traverser	83
5.4.3	Model Checker	83
5.5	ASKALON Workflow Hosting Environment	86
5.6	Real-World Scientific Workflow Modeling and Code Generation	87
5.7	Summary	89

Part III Optimization

6	Collection-Oriented Data Flow Support for Scientific Workflows	93
6.1	Introduction	93
6.2	Data Collections and Collection Distribution Constructs	94
6.2.1	Data Collections	94
6.2.2	Collection Distribution Constructs	95
6.3	Case Study	100
6.3.1	WIEN2k	100
6.3.2	MeteoAG	102
6.3.3	AstroGrid	102
6.3.4	GRASIL	104
6.4	Experimental Results	105
6.4.1	Results of the WIEN2k Workflow	106
6.4.2	Results of the MeteoAG Workflow	109
6.5	Summary	111

Part IV Synthesis

7	Semantic-Based Scientific Workflow Composition	115
7.1	Introduction	115
7.2	Ontologies	117
7.3	Ontology-Based Scientific Workflow Representation	120
7.3.1	Activity Function	120
7.3.2	Activity Type	121
7.3.3	From Semantics to Syntax	123
7.4	Implementation	127
7.5	Experimental Results	131
7.6	Summary	134

8	Automatic Scientific Workflow Composition	135
8.1	Introduction	135
8.2	Activity Function	136
8.3	Definition of the Scientific Workflow Composition Problem	137
8.4	ADD Graph and Notations	138
8.5	Control Flow Composition Algorithm	143
8.5.1	ADD Graph Creation	143
8.5.2	Workflow Extraction	143
8.5.3	Workflow Optimization	146
8.5.4	Algorithm Analysis	149
8.5.5	Branches and Loops	152
8.6	Data Flow Composition Algorithm	154
8.7	Experimental Results	156
8.8	Summary	164

Part V Related Work

9	Related Work	169
9.1	Overview of Main Workflow Systems and Languages	169
9.2	Programming	172
9.2.1	Abstraction	173
9.2.2	Type System	175
9.2.3	Control Flow	177
9.2.4	Data Flow	180
9.2.5	Modularization and Reuse	181
9.2.6	Modeling Support	182
9.2.7	Semantic Support	183
9.3	Optimization	185
9.4	Synthesis	187

Part VI Conclusions

10	Conclusions	191
10.1	Contributions	191
10.1.1	Scientific Workflow Language	191
10.1.2	Scientific Workflow Modeling	192
10.1.3	Data Flow Optimization	192
10.1.4	Semantic Scientific Workflows	193
10.1.5	Automatic Scientific Workflow Composition	194
10.2	Future Research	194

Part VII Appendices

A	Acronyms	199
----------	-----------------	-----

Contents	xiii
B Symbols	201
References	205
Index	217



<http://www.springer.com/978-3-642-30714-0>

Scientific Workflows

Programming, Optimization, and Synthesis with
ASKALON and AWDL

Qin, J.; Fahringer, Th.

2012, XXII, 222 p., Hardcover

ISBN: 978-3-642-30714-0