

Table of Contents

1. Introduction	1
1.1 Motivations	1
1.2 Problem Statement and Approach Description	3
1.2.1 General Characteristics	3
1.2.2 Derivation of Interscheme Properties	4
1.2.3 Construction of a Data Repository	8
1.2.4 Exploiting the Data Repository for Constructing a CIS or a DW	8
1.3 Related Works	10
1.3.1 Interscheme Property Derivation	10
1.3.2 Data Repository Construction	12
1.3.3 Synthesis of Cooperative Information Systems	12
1.3.4 Synthesis of Data Warehouses	14
1.4 Contribution	16
1.4.1 Interscheme Property Derivation	16
1.4.2 Data Repository Construction	17
1.4.3 Synthesis of a Cooperative Information System	18
1.4.4 Synthesis of a Data Warehouse	18
1.4.5 Experiments	19
1.5 Plan of the Thesis	21

Part I. Property Extraction

2. Extraction of Synonymies, Homonymies and Type Conflicts	27
2.1 General Characteristics of the Approach	27
2.2 Definition of the Graph-Based Metrics	29
2.3 External Fixpoint	32
2.3.1 Overview	32
2.3.2 Technical Details	34
2.4 Internal Fixpoints	44
2.4.1 Overview	44
2.4.2 Technical Details	44
2.5 A Complete Example	47

2.6	Complexity Issues	54
2.7	Sensitivity Analysis	59
2.7.1	Changes in the LSPD	59
2.7.2	Changes in the Weights $w_s(i)$	61
2.7.3	Changes in the Weight w_ζ	61
2.7.4	Changes in the Weights w_n, w_d and w_k	62
2.7.5	Result Summary	62
3.	Extraction of Object Cluster Similarities	63
3.1	General Characteristics of the Approach	63
3.2	Derivation of Object Cluster Similarities	64
3.3	A Complete Example	66
3.4	Complexity Issues	67
4.	Extraction of Hyponymies and Overlappings	71
4.1	Introduction	71
4.2	Deriving Basic Hyponymies and Overlappings	74
4.2.1	General Characteristics of the Approach	75
4.2.2	Technical Details	76
4.2.3	Tuning Thresholds and Factors	80
4.2.4	A Complete Example	80
4.3	Extracting Further Hyponymies	82
4.3.1	Description of the Approach	82
4.3.2	An Example	84
5.	Extraction of Assertions between Knowledge Patterns	87
5.1	General Characteristics of the Approach	87
5.2	The Language DL_P	89
5.2.1	Preliminary Definitions	89
5.2.2	Syntax	90
5.2.3	Semantics	91
5.3	Constructing the Inclusion Dictionary	92
5.4	Singling out the Most Interesting Classes	93
5.5	Discovering Complex Interscheme Properties	95
5.5.1	Expressions Containing \sqcap and \sqcup	96
5.5.2	Expressions Containing \exists	96
5.5.3	Expressions Containing \forall	98
5.5.4	Computing Complex Expressions	99
5.5.5	Negation	99
5.5.6	Discarding Weak or Repeated Assertions	100
5.6	A Complete Example	100
5.7	Application Domains	105
5.7.1	Query Optimization	105
5.7.2	View Materialization	107
5.7.3	Web-Based Information Systems	108

**Part II. Construction of a Cooperative Information System
and of a Data Warehouse**

6.	Construction of a Data Repository	113
6.1	Introduction	113
6.2	Support Intensional Information Base	116
6.2.1	The Metascheme	116
6.2.2	The Meta-operators	120
6.3	Constructing a Data Repository	122
6.3.1	Description of the Approach	122
6.3.2	Complexity Issues	124
6.4	Grouping Schemes into Clusters	124
6.5	Scheme Integration	126
6.5.1	Merging Entities	129
6.5.2	Handling IntrascHEME Hyponyms	130
6.6	Scheme Abstraction	133
6.6.1	Procedure Compute_AD.Rel	135
6.6.2	Procedure Compute_AD.Isa	136
6.6.3	Procedure Compute_AD.Ent	137
6.6.4	Procedure Abstract_CR	138
6.6.5	Procedure Abstract_Rel	138
6.6.6	Procedure Abstract_Isa	139
6.6.7	Procedure Abstract_Ent	140
6.7	The “Template” Language for Defining <i>IIB.M.SoV</i>	140
6.7.1	Merging Entities	140
6.7.2	Merging Attributes	141
6.7.3	Merging Relationships	141
6.7.4	Creation of an is-a Relationship	142
6.7.5	Modification of an is-a Relationship	142
6.7.6	Abstracting a Cyclic Relationship into the Corresponding Entity	142
7.	Construction of a Cooperative Information System	143
7.1	Introduction	143
7.2	General Characteristics of the Proposed Architecture	146
7.3	Structure of the Proposed CIS	148
7.4	CIS Functionalities	150
7.5	Structure of the Mediator	153
7.6	Structure of the Wrappers	155
7.7	The Web User Interface	156
8.	Construction of a Data Warehouse	161
8.1	Introduction	161
8.2	General Characteristics of the Proposed Architecture	164

8.3	Technical Details	165
8.4	Automatic Synthesis of a Data Warehouse	168

Part III. System Description and Experimentations

9.	The System D.I.K.E.	175
9.1	Introduction	175
9.2	Context Diagram of DIKE	176
9.3	Description of DIKE Modules	178
9.3.1	Intensional Knowledge Extractor (IKE)	178
9.3.2	Data Repository Constructor (DRC)	180
9.3.3	Cooperative Information System Constructor (CISC) .	181
9.3.4	Data Warehouse Constructor (DWC)	182
9.4	A Trip throughout DIKE	183
10.	Experiments on the Derivation of Similarities	
	and Type Conflicts	189
10.1	The Italian Central Governmental Office Databases	189
10.2	The Six Groups of Databases under Consideration	191
10.2.1	Group A	191
10.2.2	Group B	191
10.2.3	Group C	191
10.2.4	Group D	195
10.2.5	Group E	196
10.2.6	Group F	197
10.3	Group A – Similarity and Type Conflict Extraction	203
10.3.1	Properties Holding between the Objects of TCD and of RD	203
10.3.2	Properties Holding between the Objects of RD and MED	205
10.4	Group B – Similarity and Type Conflict Extraction	207
10.4.1	Interesting Synonymies	207
10.4.2	Interesting Homonymies	207
10.4.3	Interesting Type Conflicts	207
10.4.4	Interesting Object Cluster Similarities	207
10.5	Group C – Similarity and Type Conflict Extraction	208
10.5.1	Interesting Synonymies	208
10.5.2	Interesting Homonymies	208
10.5.3	Interesting Type Conflicts	209
10.5.4	Object Cluster Similarities	209
10.6	Group D – Similarity and Type Conflict Extraction	209
10.6.1	Interesting Synonymies	209
10.6.2	Interesting Homonymies	209
10.6.3	Interesting Type Conflicts	209

10.6.4	Interesting Object Cluster Similarities	209
10.7	Group E – Similarity and Type Conflict Extraction	212
10.7.1	Interesting Synonymies	212
10.8	Group F – Similarity and Type Conflict Extraction	212
10.8.1	Interesting Synonymies	212
10.8.2	Interesting Homonymies	213
10.8.3	Interesting Type Conflicts	213
10.8.4	Interesting Object Cluster Similarities	214
11.	Experiments on the Extraction of Hyponymies	215
11.1	Group A: Hyponymy and Overlapping Extraction	215
11.1.1	Properties Holding between Objects of RD and of MED	215
11.1.2	Properties between RD and TCD	217
11.1.3	Properties between MED and TCD	219
12.	Experiments on the Extraction of Assertions between Knowledge Patterns	221
12.1	Group E – Extraction of Assertions between Knowledge Patterns	221
13.	Experiments on the Construction of a Data Repository	225
13.1	Group A – Integration of Schemes and Abstraction of the Global Scheme	225
13.1.1	Integration Algorithm	225
13.1.2	Abstraction Algorithm	227
13.2	Group B – Integration of Schemes and Abstraction of the Global Scheme	228
13.2.1	Integration Algorithm	229
13.2.2	Abstraction Algorithm	230
13.3	Group C – Integration of Schemes and Abstraction of the Global Scheme	232
13.3.1	Integration Algorithm	232
13.3.2	Abstraction Algorithm	233
13.4	Group D – Integration of Schemes and Abstraction of the Global Scheme	234
13.4.1	Integration Algorithm	235
13.4.2	Abstraction Algorithm	236
13.5	Group F – Integration of Schemes and Abstraction of the Global Scheme	237
13.5.1	Integration Algorithm	237
13.5.2	Abstraction Algorithm	240
14.	Using the CIS Relative to ICGO Databases	243
14.1	General Options of the ICGO CIS	243

14.2 Repository Querying	244
14.2.1 Direct Access	244
14.2.2 Access Based on Schemes	246
14.2.3 Access Based on Sources	248
14.2.4 Query Formulation	248
14.3 Deferred Answer Visualization	250
<hr/>	
Part IV. Final Issues	
<hr/>	
15. A Look at the Future	255
15.1 Introduction	255
15.2 The Object Exchange Model	256
15.3 The Conceptual Model and the Related Metrics	257
15.3.1 The SDR-Network	257
15.3.2 Definition of the Metrics	263
16. Conclusions	267
References	269
A. Appendix	279
A.1 Entity-Relationship Schemes	279
A.2 Type Conflict Resolution	280
A.2.1 Resolution of an Entity-Relationship Conflict	281
A.2.2 Resolution of an Entity Attribute-Entity Conflict	282
A.2.3 Resolution of a Relationship Attribute-Entity Conflict	283
A.2.4 Resolution of an Entity Attribute (resp., a Relationship Attribute)-Relationship Conflict	283
Index	285

Extraction and Exploitation of Intensional Knowledge
from Heterogeneous Information Sources
Semi-Automatic Approaches and Tools

Ursino, D.

2002, XXVI, 292 p., Softcover

ISBN: 978-3-540-43347-7