

Contents

1	Introduction to E-Librarian Services	1
1.1	From Ancient to Digital Libraries	1
1.2	From Searching to Finding	4
1.2.1	Searching the Web	4
1.2.2	Searching Multimedia Knowledge Bases	6
1.2.3	Exploratory Search	6
1.3	E-Librarian Services	7
1.3.1	Overview	7
1.3.2	Early Question-Answering Systems	8
1.3.3	Natural Language Interface	8
1.3.4	No Library without a Librarian	9
1.3.5	Characteristics of an E-Librarian Service	10
1.4	Overview and Organization of the Book	11

Part I Key Technologies of E-Librarian Services

2	Semantic Web and Ontologies	15
2.1	What is the Semantic Web?	15
2.1.1	The Vision of the Semantic Web	15
2.1.2	Semantic Web vs. Web <i>n.0</i>	16
2.1.3	Three Principles Ruling the Semantic Web	17
2.1.4	Architecture	17
2.2	Ontologies	18
2.2.1	Ontology Structure	18
2.2.2	Upper- and Domain Ontologies	20
2.2.3	Linked Data	21
2.2.4	Expressivity of Ontologies	23
2.3	XML – Extensible Markup Language	24
2.3.1	XML: Elements, Attributes and Values	25
2.3.2	Namespaces and Qualified Names	26

2.3.3	XML Schema	26
2.3.4	Complete Example	27
2.3.5	Limitations of XML	30
2.4	RDF – Resource Description Framework	30
2.4.1	RDF Triples and Serialization	30
2.4.2	RDF Schema	32
2.4.3	Complete Example	33
2.4.4	Limitations of RDF	35
2.5	OWL 1 and OWL 2 – Web Ontology Language	36
2.5.1	Instances, Classes and Restrictions in OWL	37
2.5.2	Complete Example	38
2.5.3	From OWL 1 to OWL 2	40
2.5.4	SPARQL, the Query Language	41
3	Description Logics and Reasoning	43
3.1	DL – Description Logics	43
3.1.1	Concept Descriptions	43
3.1.2	DL Languages	44
3.1.3	Equivalences between OWL and DL	45
3.2	DL Knowledge Base	46
3.2.1	Terminologies (TBox)	46
3.2.2	World Descriptions (ABox)	48
3.3	Interpretations	48
3.3.1	Interpreting Individuals, Concepts, and Roles	48
3.3.2	Modeling the Real World	49
3.4	Inferences	51
3.4.1	Standard Inferences	52
3.4.2	Non-Standard Inferences	55
4	Natural Language Processing	61
4.1	Overview and Challenges	61
4.1.1	Syntax, Semantics and Pragmatics	61
4.1.2	Difficulties of NLP	62
4.1.3	Zipf’s law	63
4.2	Dealing with Single Words	63
4.2.1	Tokenization and Tagging	63
4.2.2	Morphology	65
4.2.3	Building Words over an Alphabet	66
4.2.4	Operations over Words	66
4.3	Semantic Knowledge Sources	67
4.3.1	Semantic relations	67
4.3.2	Semantic resources	68
4.4	Dealing with Sentences	69
4.4.1	Phrase Types	69
4.4.2	Phrase Structure	70

4.4.3	Grammar	71
4.4.4	Formal languages	72
4.4.5	Phrase structure ambiguities	72
4.4.6	Alternative parsing techniques	74
4.5	Multi-Language	75
4.6	Semantic Interpretation	77
5	Information Retrieval	81
5.1	Retrieval Process	81
5.2	Document Indexation and Weighting	82
5.2.1	Index of terms	82
5.2.2	Weighting	84
5.3	Retrieval Models	86
5.3.1	Boolean Model	87
5.3.2	Vector Model	88
5.3.3	Probabilistic Model	90
5.3.4	Page Rank	92
5.3.5	Semantic Distance	94
5.3.6	Other Models	96
5.4	Retrieval Evaluation	97
5.4.1	Precision, Recall, and Accuracy	97
Part II Design and Utilization of E-Librarian Services		
6	Ontological Approach	103
6.1	Expert Systems	103
6.1.1	Classical Expert Systems	103
6.1.2	Ontology-Driven Expert Systems	105
6.2	Towards an E-Librarian Service	106
6.2.1	Reasoning Capabilities of an E-Librarian Service	106
6.2.2	Deploying an Ontology	107
6.2.3	Designing the Ontological Background	109
6.3	Semantic Annotation of the Knowledge Base	110
6.3.1	Computer-Assisted Creation of metadata	111
6.3.2	Automatic Generation of metadata	112
7	Design of the Natural Language Processing Module	117
7.1	Overview of the Semantic Interpretation	117
7.1.1	Logical Form	117
7.1.2	Processing of a User Question	118
7.2	NLP Pre-Processing	119
7.2.1	Domain Language	119
7.2.2	Lemmatization	119
7.2.3	Handling Spelling Errors	120

7.3	Ontology Mapping	120
7.3.1	Domain Dictionary	121
7.3.2	Mapping of Words	121
7.3.3	Resolving Ambiguities	123
7.4	Generation of a DL-Concept Description	126
7.4.1	Without Syntactic Analysis	126
7.4.2	With Syntactic Analysis	127
7.4.3	How much NLP is Sufficient?	130
7.4.4	Optimization and Normal Form	130
7.5	General Limitations and Constraints	131
7.5.1	Role Quantifiers	131
7.5.2	Conjunction and Disjunction	132
7.5.3	Negation	134
7.5.4	Open-Ended and Closed-Ended Questions	135
7.5.5	Formulations	137
7.5.6	Others	138
7.6	Multiple-Language Feature	139
8	Designing the Multimedia Information Retrieval Module ..	141
8.1	Overview of the MIR Module	141
8.1.1	Knowledge Base and metadata	141
8.1.2	Retrieval Principle	143
8.1.3	The Concept Covering Problem	143
8.2	Identifying Covers	145
8.3	Computing the Best Covers	146
8.3.1	Miss and Rest	146
8.3.2	Size of a Concept Description	148
8.3.3	Best Covers	149
8.4	Ranking	150
8.5	Algorithm for the Retrieval Problem	151
8.6	User Feedback	152
8.6.1	Direct User Feedback	153
8.6.2	Collaborative Tagging and Social Networks	153
8.6.3	Diversification of User Feedback	154
9	Implementation	155
9.1	Architecture	155
9.1.1	Knowledge Layer	155
9.1.2	Inference Layer	156
9.1.3	Communication Layer	157
9.1.4	Presentation Layer	157
9.2	Development Details	158
9.2.1	Processing OWL and DL in Java	158
9.2.2	Client Front-End with Ajax Autocompleter	162
9.2.3	The SOAP Web Service Interface	163

Part III Applications

10 Best practices	167
10.1 Computer History Expert System (CHESt)	167
10.1.1 Description	167
10.1.2 Experiment	169
10.2 Mathematics Expert System (MatES)	170
10.2.1 Description	170
10.2.2 Benchmark Test	171
10.2.3 Experiment	174
10.3 The Lecture Butler's E-Librarian Service	175
10.3.1 Description	175
10.3.2 Benchmark Tests	176

Part IV Appendix

A XML Schema Primitive Datatypes	183
B Reasoning Algorithms	185
B.1 Overview	185
B.2 Structural Subsumption	185
B.2.1 Example 1	186
B.2.2 Example 2	186
C Brown Tag Set	187
D Part-of-Speech Taggers and Parsers	191
D.1 POS Taggers	191
D.2 Parsers	192
E Probabilistic IR Model	193
E.1 Probability Theory	193
E.2 Probabilistic Model	194
References	197
Index	205