

TABLE OF CONTENTS

	Page
TITLE PAGE	i
APPROVAL PAGE	ii
DECLARATION	iii
ACKNOWLEDGEMENTS.....	iv
TABLE OF CONTENTS	v
LIST OF TABLES.....	viii
LIST OF FIGURES.....	ix
LIST OF TERMS AND ABBREVIATIONS.....	xi
INTISARI	xiii
ABSTRACT.....	xiv
CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Problem Formulation	5
1.3 Research Assumptions	6
1.4 Research Goals.....	7
1.5 Research Benefits.....	8
1.6 Research Contributions.....	9
CHAPTER II LITERATURE REVIEW	11
2.1 Question Analysis.....	12
2.1.1 Text-based why-question analysis	12
2.1.2 Question analysis in ontology-based QA.....	15
2.1.3 Querying in ontology-based IR.....	18
2.2 Document Retrieval	21
2.2.1 Document Retrieval of Why-Question Answering	21
2.2.2 Ontology-Based Document Retrieval.....	22
2.3 Answer Construction.....	23
2.3.1 Answer Extraction of Why-Question Answering.....	23
2.3.2 Text-Based KI.....	26
2.4 Summary	29
CHAPTER III THEORETICAL BACKGROUND	30
3.1 Definitions	30
3.2 Description Logic.....	32
3.3 Ontology.....	33
3.3.1 Basic of ontology	33
3.3.2 Domain ontology.....	33
3.3.3 Description logics as ontology languages.....	33
3.3.4 Ontology learning.....	34
3.4 Text-Based Why-Question Answering	34
3.4.1 Characteristic of natural language question	34
3.4.2 Expected answer type of a why-question	34
3.4.3 Why-question answering phases	35
3.5 Information Retrieval.....	37
3.5.1 Vector space model.....	37
3.5.2 Ontology-based IR.....	39

3.6	Stanford typed dependency grammar	42
3.7	Standard Evaluation Measures for Question Answering System	42
3.8	Evaluation Measures for Semantic Representation	43
CHAPTER IV RESEARCH METHODOLOGY		45
4.1	Architecture of the Proposed Method.....	45
4.2	The why-question analysis.....	47
4.3	The document retrieval	48
4.4	The answer construction	48
4.4.1	The answer extraction.....	48
4.4.2	The answer integration	49
4.5	Domain ontology building.....	50
4.6	Data.....	51
4.6.1	Question collection	51
4.6.2	Document corpus	54
4.7	The multi-word term extraction patterns, the triple construction patterns, the SPARQL templates, and the answer interpretation templates.....	55
4.8	Evaluation	56
4.8.1	Experiments.....	56
4.8.2	Datasets.....	57
4.8.3	Evaluation metrics.....	58
4.9	Tools.....	59
CHAPTER V ONTOLOGY-BASED WHY-QUESTION ANALYSIS METHOD		61
5.1	The Proposed Why-Question Analysis Method	61
5.2	The Query Formulation.....	64
5.2.1	The proposed information extraction method	64
5.2.2	The semantic entity mapping method	77
5.3	The Query Reformulation Method	80
5.4.1	The causality annotation identification.....	80
5.4.2	The original semantic annotation identification	82
5.4.3	The query expansion for identifying the additional semantic annotations	82
5.4	Evaluation of the Query Formulation and Query Reformulation	87
5.4.1	Experiments and data	88
5.4.2	Results and discussions	89
5.5	Evaluation of the Proposed Why-Question Analysis in Document Retrieval	91
5.5.1	Method Description	91
5.5.2	Evaluation data and metrics.....	93
5.5.3	Results and discussions	94
5.6	Conclusions	96
CHAPTER VI ONTOLOGY-BASED DOCUMENT RETRIEVAL METHOD		98
6.1	The Proposed Document Retrieval Method	98
6.2	The Semantic Index Construction	101
6.2.1	The dataset of label and label-instance pair construction.....	103
6.2.2	The one-word term index construction	103
6.2.3	The multi-word term index construction.....	104
6.2.4	The multi-word term index reconstruction.....	105
6.2.5	The semantic index construction	106
6.2.6	The inverted semantic index construction.....	107

6.2.7	The <i>TFIDF</i> matrix construction	107
6.3	The Searching Method	108
6.4	The Proposed Ranking Method	109
6.4.1	Phase 1: <i>TFIDF</i> -based similarity	109
6.4.2	Phase 2: Causality-detection-based similarity	112
6.4.3	Phase 3: Linear combination of <i>TFIDF</i> - and causality-detection-based similarity	123
6.5	Evaluation	124
6.5.1	Experiments and data	124
6.5.2	Results and discussions	125
6.6	Conclusions	132
CHAPTER VII ONTOLOGY-BASED ANSWER CONSTRUCTION METHOD		134
7.1	The Proposed Ontology-Based Answer Construction Method	134
7.2	The Proposed Ontology-Based Answer Extraction Method	136
7.2.1.	The paragraph filtering and indexing	138
7.2.2.	The proposed paragraph extraction	140
7.2.3.	The proposed sentence extraction	147
7.3	The Proposed Ontology-Based Answer Integration Method	152
7.3.1	The Proposed Ontology-Based Answer Integration Method	152
7.3.2	The proposed information extraction method	154
7.3.3	The semantic entity mapping method	164
7.3.4	The proposed semantic-triples integration	165
7.3.5	The proposed integrated semantic-triples interpretation method	173
7.4	Evaluation of the Proposed Answer Extraction Method	189
7.4.1	Method description	190
7.4.2	Evaluation data and metrics	192
7.4.3	Results and Discussions	193
7.5	Evaluation of the Proposed Answer Integration Method	199
7.6	Conclusions	202
CHAPTER VIII CONCLUSION		204
8.1	Conclusions	204
8.2	Future Works	206
8.2.1	Improving the method performance	206
8.2.2	Expansion to other languages	207
REFERENCES		208
APPENDIX A Patterns for Extracting the Multi-word Terms from Text		214
APPENDIX B Patterns for Constructing Semantic Representation of a Why-Question ..		215
APPENDIX C Constructing Semantic Representation of the Causal Sentences		220
APPENDIX D Interpreting the Label-Based Triples		228
APPENDIX E Examples of Datasets for Evaluations		232
APPENDIX F Expansion to other languages		238
F.1.	Question analysis for non-English question with English corpus available	238
F.2.	The why QA approach for non-English corpus	240