

TABLE OF CONTENTS

LIST OF FIGURES	10
LIST OF TABLES	11
ABSTRACT	12
1 INTRODUCTION	14
2 LITERATURE REVIEW	17
2.1 Static Nature of Knowledge in Library Collection	17
2.1.1 Characterizing Document Objects	18
2.1.1.1 Theory of Indexing	18
2.1.1.2 Manual Indexing	22
2.1.1.3 Indexing with Controlled Vocabulary and Thesauri	24
2.1.1.4 Free-text Indexing	25
2.1.2 Characterizing Global Knowledge in Document Collections	27
2.1.2.1 Classification Schemes and Categorization	27
2.1.2.2 Knowledge Discovery from Large Databases	30
2.1.2.3 Knowledge Bases, Inferencing, and Ontology	32
2.1.3 Revealing Knowledge in Neighborhood	34
2.1.3.1 Syntactic Mapping: Index List	34
2.1.3.2 Keyword Mapping: Controlled Vocabulary	35
2.1.3.3 Semantic Mapping: Man-made Thesauri	36
2.1.3.4 Co-occurred Mapping: Automatic Thesauri	37
2.2 Dynamic Nature of User Information Need	38
2.2.1 Expressing User Need	39
2.2.1.1 Information Need	39
2.2.1.2 Indeterminism	41
2.2.1.3 Opportunism	42
2.2.1.4 Vocabulary Problem	43
2.2.1.5 Recognition with Contextual Information	44
2.2.2 Perceiving Knowledge	45
2.2.2.1 Computing Relevance?	48
2.2.2.2 Structureless and Contextless: Document List	50
2.2.2.3 Structural but Contextless: Dynamic Clustering	51
2.2.2.4 Sturctural and Contextual: Path to the Knowledge	52

TABLE OF CONTENTS – *Continued*

3	RESEARCH QUESTIONS AND METHODOLOGIES	56
3.1	Research Questions	56
3.2	Methodologies	58
3.3	Dissertation Plan	60
3.3.1	A Framework for Semantic-enabled Information Technologies	60
3.3.2	Structure of Dissertation	67
4	CONCEPT SPACE CONSULTATION	69
4.1	Objectives	69
4.2	Research Questions and Methodology	71
4.3	Background and Issues	72
4.3.1	Spreading Activation in a Concept Network	72
4.3.2	Semantic Nets	73
4.3.3	Neural Nets	75
4.4	Concept Exploration in a Large Text-based Concept Network	78
4.4.1	Databases	79
4.4.2	Knowledge Discovery Methods	80
4.4.3	Knowledge Sources for Concept Network	82
4.4.4	Concept Exploration Methods	88
4.5	Two Algorithms for Spreading Activation	92
4.5.1	A Branch-and-bound Spreading Activation Algorithm: Semantic Net Based	92
4.5.2	A Hopfield Net Spreading Activation Algorithm: Neural Net Based	97
4.6	System Evaluation: A Benchmark Testing	102
4.7	User Evaluation: A Concept Exploration Experiment	103
4.7.1	Experimental Design	103
4.7.2	Experimental Results: Performance Analysis on Concept Retrieval	106
4.7.3	Experimental Results: Performance Analysis on Document Retrieval	114
4.8	Discussion and Conclusion	118

TABLE OF CONTENTS – *Continued*

5	LARGE-SCALE CONCEPT SPACE GENERATION	122
5.1	Objectives	122
5.2	Research Questions and Methodology	123
5.3	Background and Issues	124
5.3.1	Alleviating the Vocabulary Problem Using Concept Spaces .	124
5.3.2	Concept Association and Thesaurus Work	125
5.4	Concept Space Techniques	128
5.4.1	Document and object list collection	129
5.4.2	Object Filtering	130
5.4.3	Automatic Indexing	130
5.4.4	Co-occurrence Analysis	131
5.4.5	Associative Retrieval	134
5.5	Parallel Computing Approach	134
5.5.1	Parallel Computing for Information Retrieval	134
5.5.2	Parallel Computing for Knowledge Discovery	137
5.6	Parallel Computing: Concept Space Generation	139
5.6.1	Worm and Fly Concept Spaces Generation Using CM-5 and SGI Power Challenge	140
5.6.2	Computer Engineering Concept Space Generation Using SGI Power Challenge	144
5.6.3	CancerLit Concept Space Generation Using SGI Origin2000	146
5.6.4	Lessons Learned: Matching Concept Space Techniques with Parallel Computer Architecture	149
5.6.4.1	Data Parallelization	150
5.6.4.2	Algorithmic Parallelization	150
5.6.4.3	Synchronization: How Often?	152
5.7	System Evaluation: A Concept Association Experiment . .	152
5.7.1	Experimental Design	153
5.7.2	Experimental Results: Concept Recall and Concept Precision	154
5.8	Discussion and Conclusion	159

TABLE OF CONTENTS – *Continued*

6 CONCLUSIONS	162
6.1 Contributions	162
6.1.1 Dynamic Use of Knowledge Sources	162
6.1.2 Complementary to Man-made Thesauri	163
6.1.3 Context-specific Concept Spaces	164
6.1.4 Concept Space As Semantic Exchange Medium	166
6.2 Lessons Learned	168
6.3 Future Directions	170
Appendix A. BENCHMARK TESTING	173
A.1 Benchmark Testing Design	173
A.2 Benchmark Testing Results and Discussion	173
Appendix B. SAMPLE SESSIONS	177
Appendix C. CANCER SPACE: A WEB-BASED INFORMATION RETRIEVAL SYSTEM	182
C.1 Client User Interface	183
C.2 Server Implementation	186
Appendix D. FUNDING AND ACKNOWLEDGMENTS	189
D.1 Concept Space Consultation	189
D.2 Concept Space Generation	189
REFERENCES	191