



















## LIST OF FIGURES

FIGURE NO.	TITLE	PAGE NO.
1.1	General Schema of IR	4
1.2	Steps in Document Classification	6
1.3	Vector Space Model for Information Retrieval	7
1.4	k-Nearest Neighbors of Classes G1 and G2	11
1.5	Structure of Artificial Neural Networks	14
3.1	Workflow Model for Semantic Feature Extraction and Indexing for TC	49
3.2	Hypernyms of two objects : Chair and Honeypot	56
3.3	Hypernym - Hyponym Relationship Representing the Semantics	57
3.4	Semantic Distance between the Hypernyms	59
3.5	Hypernym Radix Trie for Efficient Retrieval	60
3.6	Algorithm for k-NN Classification	62
3.7	Pseudo Code of Naive Bayes Algorithm	63
3.8	Mapping of testing Document to Corresponding Topic	64
3.9	Sample Training Document of "WebKB" Dataset for the topic "Faculty"	66
3.10	Selected Terms from the Training Documents	67
3.11	Classification Accuracy	69
3.12	Comparison of Precision for the Topic "Course" Using different Methods	70
3.13	Comparison of Precision for the Topic "Faculty" Using different Methods	71
3.14	Comparison of Precision for the Topic "Project" Using different Methods	71

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
3.15	Comparison of Precision for the Topic "Student" Using different Methods	72
3.16	Comparison of Recall for the Topic "Course" Using different Methods	73
3.17	Comparison of Recall for the Topic "Faculty" Using different Method	74
3.18	Comparison of Recall for the Topic "Project" Using different Methods	75
3.19	Comparison of Recall for the Topic "Student" Using different Methods	75
4.1	Processing steps in the Enhanced Topic Classification Model	79
4.2	Representation of Phrase Set, Document Set and Word Set	86
4.3	Document-Term Frequency Matrix	90
4.4	Parameters of DBN	96
4.5	Pre-Training and Fine-Tuning Stages of RBM	97
4.6	Layer-to-Layer Conditional Probability	98
4.7	Pseudo Code of RBM Update method	100
4.8	Pseudo code of SDL Training Using RBM_Update	101
4.9	Pseudo Code of RBM Fine-Tuning	103
4.10	The Layered Structure of the Proposed Semantic Deep Learner(SDL)	104
4.11	Sample Document of "20NewsGroups" Dataset Taken for Training	109

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
4.12	Semantic Smoothing Values of Terms in the Input Document	110
4.13	Experimental Results Showing Top Terms, Topics and Testing Documents of "20NewsGroups" Dataset	110
4.14	Output Values Generated by the Proposed SDL for each Topic	111
4.15	Calculated Accuracy based on the Actual Topic and Retrieved Topic	111
4.16	Performance Comparison by varying Learning Rate using '20 Newsgroups' Dataset	113
4.17	Performance Comparison by varying Learning Rate using 'WebKB' Dataset	114
4.18	Performance Comparison by varying hidden layers using '20NewsGroups' Dataset	115
4.19	Performance Comparison by varying hidden layers using 'WebKB' Dataset	116
4.20	Performance Comparison by varying epoch using '20NewsGroups' Dataset	117
4.21	Performance Comparison by varying epoch using 'WebKB' Dataset	117
4.22	Performance Comparison by varying Learning Rate using '20NewsGroups' Dataset (Top twenty key terms)	118
4.23	Performance Comparison by varying Learning Rate using 'WebKB' Dataset (Top twenty key terms)	119
4.24	Performance Comparison by varying hidden layers using '20Newsgroups' Dataset (Top twenty Key terms)	120

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
4.25	Performance Comparison by varying hidden layers using 'WebKB' Dataset (Top twenty key terms)	120
4.26	Performance Comparison by varying epoch using '20NewsGroups' Dataset (Top twenty key terms)	121
4.27	Performance Comparison by varying epoch using 'WebKB' Dataset (Top twenty key terms)	122
5.1	An Intelligent Topic Classification Model with enriched semantic features	127
5.2	Rule-based decision tree for selection of sense	133
5.3	Document-Term Training Matrix for Selected Significant terms	134
5.4	The Conceptual Schema of the Proposed Intelligent Topic Classification Model	136
5.5	Sample Document Under the Topic "Voltage Stability"	138
5.6	Portion of the Decision List and Top Terms Generated Through DLWSD	139
5.7	Sample Features with Semantic Score Using 'Electrical' Dataset	140
5.8	Top Ten Terms, Topics and Testing Documents with Respect to 'Electrical' Dataset	141
5.9	Output Values Generated by ESDL using 'Electrical' Dataset	141
5.10	The Sample Output Showing the Evaluated Accuracy based on Actual_ID and Retrieved_ID	142
5.11	Performance Comparison by Varying Learning Rate Using '20NewsGroups' Dataset	144

<b>FIGURE NO.</b>	<b>TITLE</b>	<b>PAGE NO.</b>
5.12	Performance Comparison by Varying Learning Rate Using 'WebKB' Dataset	145
5.13	Performance Comparison by Varying Learning Rate Using 'Electrical' Dataset	145
5.14	Performance Comparison by Varying Hidden Layers Using '20NewsGroups' Dataset	146
5.15	Performance Comparison by Varying Hidden Layers Using 'WebKB' Dataset	147
5.16	Performance Comparison by Varying Hidden Layers Using 'Electrical' Dataset	147
5.17	Performance Comparison by Varying Epoch Using '20NewsGroups' Dataset	148
5.18	Performance Comparison by Varying Epoch Using 'WebKB' Dataset	149
5.19	Performance Comparison by Varying Epoch Using 'Electrical' Dataset	149
6.1	Sequential Steps in the Proposed Patent Classification	159
6.2	Classification Accuracy of SDL for the Predicted Topics	163
6.3	Accuracy Comparison of SDL with Neural Network based Classifier on Different Topics of Patent Dataset	164

