

# Contents

<b>Abstract</b>	<b>iii</b>
<b>Declaration</b>	<b>v</b>
<b>Acknowledgment</b>	<b>ix</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Introduction . . . . .	1
1.1.1 Definition of Named Entity Recognition . . . . .	2
1.1.2 Application of Named Entity Recognition . . . . .	3
1.2 Problems in Named Entity Recognition . . . . .	4
1.3 Approaches to Named Entity Recognition . . . . .	9
1.4 Conclusion . . . . .	19
<b>2 NER Research in Indian Languages</b>	<b>21</b>

2.1	Introduction . . . . .	21
2.2	NER work in Indian languages . . . . .	22
2.2.1	NER in Bengali . . . . .	23
2.2.2	NER In Hindi . . . . .	28
2.2.3	NER in Telugu . . . . .	31
2.2.4	NER in Tamil . . . . .	32
2.2.5	NER In Oriya . . . . .	33
2.2.6	NER in Urdu . . . . .	34
2.2.7	NER in Punjabi . . . . .	35
2.2.8	NER in Kannada . . . . .	36
2.2.9	International Work done on NER . . . . .	37
2.2.10	Work done on NER using Gazetteer . . . . .	39
2.3	Tagsets . . . . .	42
2.3.1	The IJCNLP-08 Shared Task Tagset . . . . .	42
2.3.2	The Named Entity Tagset by Ekbal and Bandopadhyay . . . . .	42
2.3.3	CoNLL 2003 Shared Tagset . . . . .	43
2.4	Evaluation Metrics used in NER . . . . .	43
2.5	Cross-validation . . . . .	45

2.6	Conclusion . . . . .	46
<b>3</b>	<b>Rule-based and Gazetteer-based NER for Assamese</b>	<b>49</b>
3.1	Introduction . . . . .	49
3.2	Rule-based Approach . . . . .	50
3.2.1	Different Approaches to Stemming . . . . .	50
3.2.2	Previous Work on Stemming . . . . .	53
3.2.3	Assamese Language and its Morphology . . . . .	54
3.2.4	Assamese Corpora . . . . .	57
3.2.5	Our Approach . . . . .	58
3.2.6	Results and Discussions . . . . .	61
3.3	Gazetteer-based Approach . . . . .	68
3.3.1	Results and Discussions . . . . .	72
3.4	Conclusion . . . . .	73
<b>4</b>	<b>Machine Learning based NER for Assamese</b>	<b>74</b>
4.1	Introduction . . . . .	74
4.2	Features used in Named Entity Recognition . . . . .	75
4.3	CRF Approach . . . . .	78

4.3.1	Experiment . . . . .	79
4.3.2	Results and Discussion . . . . .	83
4.4	The HMM Approach . . . . .	86
4.4.1	Our Experiment . . . . .	88
4.4.2	Results and Discussion . . . . .	91
4.5	A Hybrid approach for NER . . . . .	97
4.5.1	Previous work On NER using Hybrid Approach . . . . .	102
4.5.2	Results Discussions . . . . .	107
4.6	Conclusion . . . . .	108
<b>5</b>	<b>Conclusion and Future Directions</b>	<b>109</b>
	<b>Publications</b>	<b>112</b>
<b>A</b>	<b>Gazetteer List</b>	<b>115</b>