

List of Tables

2.1	NER approaches in Indian languages	47
2.2	NER approaches in Indian languages	48
3.1	Example of suffixes found in locations and organizations	56
3.2	Examples of location named entities	58
3.3	Suffix stripping experiment for location named entity	61
3.4	Examples of issues in preparing gazetteer list	69
3.5	Data for gazetteer list	70
3.6	Results obtained using gazetteer list	71
4.1	Example of training file for CRF	81
4.2	Format of the Serialized file	83
4.3	NER results for Set 1 using CRF	84
4.4	NER results for Set 2 using CRF	84

4.5	NER results for Set 3 using CRF	85
4.6	Average CRF Results	85
4.7	NER results for Set 1 using HMM	91
4.8	NER results for Set 2 using HMM	92
4.9	NER results for Set 3 using HMM	92
4.10	Average HMM Results	92
4.11	NER results for Set 1 using HMM and Smoothing	95
4.12	NER results for Set 2 using HMM and Smoothing	95
4.13	NER results for Set 3 using HMM and Smoothing	95
4.14	Average HMM Results after applying Smoothing	96
4.15	NER results for Bodo Dataset	97
4.16	NER results for BPM Dataset	97
4.17	NER results for Set 1 using Hybrid Approach	99
4.18	NER results for Set 2 using Hybrid Approach	99
4.19	NER results for Set 3 using Hybrid Approach	102
4.20	Average Result for Hybrid Approach	102
4.21	Different work on NER using Hybrid approach	103
4.22	Different work on NER in Bengali and Hindi using Hybrid approach .	104

4.23 Results obtained using gazetteer list	105
4.24 Average HMM Results	105
4.25 Average Hybrid Results	105
4.26 Results for Assamese using hybrid approach	106
4.27 NER Results fo Bengali Dataset using an hybrid approach	107
4.28 NER Results for Hindi Dataset using an hybrid approach	107

List of Figures

2.1	Comparison of approaches for different languages	41
3.1	Flow Chart Of The Process	60
4.1	State Transition Diagram for states S1,S2,S3,S4 S5	88
4.2	Hybrid NER Architecture	100
4.3	Comparison of NER for Assamese using different approaches.	101

List Of Abbreviations

AS: Automatic Summarization
CRF: Conditional Random Fields
HMM: Hidden Markov Model
IE: Information Extraction
IAA: Inter-Annotator Agreement
IR: Information Retrieval
KA: Knowledge Acquisition
ML: Machine Learning.
ME: Maximum Entropy
MEMM: Maximum Entropy Markov Model
MUC: Message Understanding Conference
MUC6: The Sixth Message Understanding Conference
NER: Named Entity Recognition
NE: Named Entity
NLP: Natural Language Processing
QA: Question Answering
SVM: Support Vector Machine