

Table of Contents

1	Introduction	1
1.1	Problem Statement	3
1.2	Hypothesis	4
1.3	Objectives	4
1.3.1	General Objective	4
1.3.2	Specific Objectives	4
1.4	Methodology	4
1.5	Thesis Structure	5
2	Background and Related Work	6
2.1	Scientific Disciplines	6
2.1.1	Artificial Intelligence	6
2.1.2	Machine Learning	7
2.1.3	Natural Language Processing	9
2.2	Named Entity Recognition	10
2.2.1	Nested Named Entity Recognition	12
2.3	Related Work	14
2.3.1	Annotated corpora	14
2.3.2	Named Entity Recognition	15
2.3.3	Nested Named Entity Recognition	16
3	Flat Named Entity Recognition in the Chilean Waiting List	19

3.1	Data Description	19
3.1.1	Annotation Process	19
3.1.2	Data Exploration	23
3.2	Methods	26
3.2.1	Embedding Layer	27
3.2.2	Encoder Layer	28
3.2.3	Classification Layer	29
3.2.4	Experimental Settings	30
3.3	Results on flat NER	32
3.4	Discussion	33
4	Nested Named Entity Recognition in the Chilean Waiting List	36
4.1	Nested NER Architectures	36
4.1.1	Multiple LSTM-CRF (MLC)	36
4.1.2	Sequence Multi-Labeling (SML)	37
4.2	Methods	38
4.2.1	Baseline	38
4.2.2	Word Representation	39
4.2.3	Settings	41
4.2.4	Model Evaluation	42
4.2.5	Error Analysis	43
4.3	Results on nested NER	43
4.3.1	Main Results	43
4.3.2	Hypothesis Test Results	45
4.3.3	Error Analysis Results	45
4.4	Demo of our Medical Entity Recognition Model	47
4.5	Discussion	47

5	Nested Named Entity Recognition Revisited	49
5.1	Motivation	49
5.2	Datasets	50
5.2.1	GENIA	51
5.2.2	GermEval	51
5.3	Methods	52
5.3.1	Baselines	52
5.3.2	Implementation Details	56
5.3.3	Evaluation Metrics	57
5.4	Results	58
5.4.1	Main Results	58
5.4.2	Nested Results	60
6	Conclusions and Future Work	63
6.1	Conclusions	63
6.2	Future Work	64
6.3	Contributions	64
	Bibliography	65