

Table of Content

1	Introduction	1
1.1	Research Problem	2
1.2	Hypothesis	4
1.3	Research Questions	4
1.4	Objectives	4
1.4.1	General Objective	4
1.4.2	Specific Objectives	5
1.5	Results	5
1.6	Research Outcome	5
1.7	Thesis Structure	6
2	Background and Related Work	7
2.1	Scientific Disciplines	7
2.1.1	Artificial Intelligence	7
2.1.2	Machine Learning	8
2.1.3	Representation Learning	8
2.1.4	Transfer Learning	9
2.1.5	Natural Language Processing	9
2.1.6	Representations of Text	11
2.2	Transformers	11
2.2.1	Vanilla Transformers	12

2.2.2	BERT	14
2.2.3	RoBERTa	14
2.2.4	ALBERT	16
2.2.5	Multilingual and Monolingual Models	17
2.3	Model Compression and Acceleration	18
2.3.1	Compression Techniques	19
2.3.2	Knowledge Distillation	20
2.3.3	Knowledge Distillation Techniques for BERT models	21
3	Preliminaries	22
3.1	Evaluation Tasks and Metrics	22
3.1.1	Evaluation Metrics	23
3.1.2	Evaluation Tasks	24
3.2	Overview of Pre-trained Models for Spanish	26
3.2.1	BETO	26
3.2.2	DistilBETO	27
3.2.3	RoBERTa-BNE	27
3.2.4	BERTIN	27
3.2.5	Models Not Included	27
4	Proposed Spanish NLP Resources	29
4.1	ALBETO: Lighter Models for Spanish	29
4.1.1	Data Collection and Preprocessing	29
4.1.2	Model Architecture	31
4.1.3	Pre-training Procedure	31
4.1.4	Fine-tuning and Experimental Setup	32
4.2	Speedy Gonzales: Faster Models for Spanish	32
4.2.1	Approach	33

4.2.2	Student Models	34
4.2.3	Applying Knowledge Distillation to Different NLP Tasks	35
4.2.4	Additional Implementation Details and Experimental Setup	37
5	Results and Discussion	40
5.1	Task Performance	40
5.2	Model Efficiency and Inference Speed	43
5.3	Inference Speed on Common Hardware	46
6	Conclusion	48
6.1	Summary of Contributions	48
6.2	Limitations and Future Research Directions	49
	Bibliography	62
	Annexes	63
	Annex A Examples of Considered Tasks	63
	Annex B Training Loss on ALBETO models	66
	Annex C Selected Teacher Models	68
	Annex D Knowledge Distillation Implementation for Models With Shared Vocabulary	69
	Annex E Effect of Caching Teacher Outputs During Training	71
	Annex F Importance of Knowledge Distillation	73