

# Tabla de Contenido

<b>1. Introduction</b>	<b>1</b>
1.1. Motivation . . . . .	1
1.2. Hypothesis . . . . .	5
1.3. Objectives . . . . .	6
1.4. Methodology . . . . .	6
1.5. Contributions . . . . .	7
1.6. Work Structure . . . . .	9
<b>2. Theoretical Framework</b>	<b>11</b>
2.1. Semantic Web . . . . .	11
2.1.1. Web of Data . . . . .	11
2.1.2. Resource Description Framework . . . . .	13
2.1.3. SPARQL Query Language . . . . .	16
2.1.4. Linked Open Data Cloud . . . . .	18
2.1.5. Wikidata . . . . .	19
2.2. Information Extraction . . . . .	22
2.2.1. Information Extraction methods with Semantic Web technologies . . . . .	22
2.2.2. Entity Linking . . . . .	23
2.2.2.1. DBpedia Spotlight . . . . .	26
2.2.2.2. AIDA . . . . .	27
2.2.2.3. TAGME . . . . .	29
2.2.2.4. OpenTapioca . . . . .	30
2.2.3. Sequence Labeling . . . . .	31
2.2.3.1. Contextual String Embeddings . . . . .	32
2.2.3.1.1. Language Models . . . . .	33
2.2.3.1.2. Extracting Flair Embeddings . . . . .	33
2.2.3.2. Sequence Labeling Architecture . . . . .	34
2.3. Semantic Parsing . . . . .	36
2.3.1. Sequence to Sequence models . . . . .	37
2.3.1.1. Convolutional Sequence to Sequence Model . . . . .	38
2.3.2. Natural Language to SPARQL . . . . .	41

2.3.2.1.	Neural Machine Translation . . . . .	42
2.4.	Question Answering over Knowledge Graphs . . . . .	44
2.4.1.	KGQA approaches . . . . .	45
2.4.1.1.	Information Retrieval-based methods . . . . .	46
2.4.1.2.	Neural Semantic Parsing-based methods . . . . .	46
2.4.2.	Main Challenges . . . . .	47
2.4.3.	Benchmark & Datasets . . . . .	49
2.4.3.1.	QALD . . . . .	49
2.4.3.2.	LC-QuAD 1 . . . . .	50
2.4.3.3.	DBNQA . . . . .	50
2.4.3.4.	LC-QuAD 2 . . . . .	51
<b>3.</b>	<b>System Overview</b>	<b>54</b>
3.1.	Question Answering general overview . . . . .	54
3.2.	Query Generation Module . . . . .	55
3.2.1.	Query Generation pipeline . . . . .	56
3.2.2.	Query Encoding . . . . .	57
3.2.3.	Fairseq Model . . . . .	58
3.3.	Entity Linking Module . . . . .	59
3.3.1.	Entity Linking pipeline . . . . .	60
3.3.2.	Individual Entity Linking systems . . . . .	61
3.3.3.	Ensemble Entity Linking system . . . . .	62
3.3.3.1.	Precision Priority system . . . . .	63
3.3.3.2.	Voting system . . . . .	64
3.3.3.3.	Other optimizations . . . . .	65
3.4.	Slot Filling Module . . . . .	66
3.4.1.	Slot Filling pipeline . . . . .	66
3.4.2.	Sequence Tagger model . . . . .	67
3.4.3.	Slot Filling method . . . . .	68
<b>4.</b>	<b>Experimental Design</b>	<b>72</b>
4.1.	Question Answering general overview . . . . .	72
4.2.	Question Answering Dataset . . . . .	73
4.2.1.	LC-QuAD 2 Dataset Cleaning . . . . .	73
4.2.2.	Query Template Dataset . . . . .	75
4.2.3.	Sequence Labeling Dataset . . . . .	76
4.2.4.	Final Dataset Format . . . . .	76
4.3.	System Implementation . . . . .	77
4.4.	Experiments . . . . .	78
4.4.1.	Datasets . . . . .	78
4.4.2.	Query Template Generation . . . . .	79
4.4.3.	Entity Linking . . . . .	79
4.4.4.	Sequence Labeling and Slot Filling . . . . .	80

4.4.5. Question Answering over Knowledge Graphs . . . . .	80
<b>5. Results</b>	<b>82</b>
5.1. Query Template generation . . . . .	82
5.2. Entity Linking . . . . .	84
5.3. Sequence Labeling and Slot Filling . . . . .	87
5.4. SPARQL Query Generation . . . . .	88
5.5. Question Answering over Knowledge Graphs . . . . .	92
<b>6. Conclusions</b>	<b>95</b>
6.1. Relevance and Contributions . . . . .	96
6.2. Future work . . . . .	97
<b>Bibliography</b>	<b>99</b>
<b>Appendix A. Neural Networks overview</b>	<b>123</b>
A.1. Neural Networks Fundamentals . . . . .	123
A.1.1. Multilayer Perceptrons . . . . .	123
A.1.2. Network Training . . . . .	125
A.2. Recurrent Neural Networks . . . . .	128
A.2.1. Long Short-Term Memory . . . . .	129
A.3. Convolutional Neural Networks . . . . .	131
A.3.1. Convolutional layer . . . . .	132
A.3.2. Pooling layer . . . . .	133
A.3.3. Common architectures . . . . .	134
<b>Appendix B. Question Answering Dataset</b>	<b>135</b>
B.1. Normalized Dataset Format . . . . .	135
B.2. LC-QuAD 2 base templates . . . . .	136