Contents

Abstract **ii**

1 Introduction **1**
1.1 Objectives . . . 2
1.1.1 Main objective . . . 2
1.1.2 Specific objectives . . . 2
1.2 Methodology . . . 2
1.3 Expected Results . . . 3
1.4 Outline of this Document . . . 4

2 Background **5**
2.1 Semantic Web . . . 5
2.1.1 Data Model . . . 6
2.1.2 Queries on RDF . . . 9
2.1.3 Linked Data . . . 9
2.1.4 Wikidata . . . 10
2.1.5 Query Interfaces . . . 12
2.2 Information Retrieval . . . 15
2.2.1 Inverted Index . . . 15
2.2.2 Ranking Importance . . . 15

3 System Overview **17**
3.1 User Interactions . . . 17
3.1.1 Keyword . . . 19
3.1.2 Type . . . 19
3.1.3 Results . . . 19
3.1.4 Facets . . . 20
3.2 Query Expressiveness . . . 21
3.2.1 Operations . . . 21
3.2.2 Analysis versus SPARQL Queries . . . 22

4 Back end **24**
4.1 Initial index . . . 24
4.1.1 Structure of the documents . . . 24
4.1.2 Building the index . . . 26
4.2 Ranking data . . . 27
4.3 Types index . . . 28
4.4 Queries and results . . . 28
4.4.1 Queries by identifier . . . 29
4.4.2 Queries by keyword . . . 29
4.4.3 Queries by type . . . 29
4.4.4 Queries by property-value . . . 30
4.4.5 Mixed queries . . . 30
4.5 Improving times and caching . . . 31
4.6 Cache indexes . . . 34
4.6.1 Cache for properties . . . 34
4.6.2 Cache for values . . . 34

5 Front end **36**
5.1 Home page . . . 36
5.2 Results page . . . 38
5.2.1 Navigation bar . . . 39
5.2.2 Results entries . . . 39
5.2.3 Facet selection . . . 40

6 Evaluation **41**
6.1 System Performance . . . 41
6.1.1 Indexes statistics . . . 41
6.1.2 Browsing performance . . . 43
6.2 User study . . . 44
6.3 Wikidata's collaborators study . . . 48
6.4 Discussion . . . 50

7 Conclusion **51**
7.1 Current limitations . . . 51
7.2 Future work . . . 52

Bibliography **54**