Table of contents

1 Introduction **1**
- 1.1 Research Hypothesis . . . 3
- 1.2 Objectives . . . 3
  - 1.2.1 General Objective . . . 3
  - 1.2.2 Specific Objectives . . . 3
- 1.3 Contributions . . . 4
- 1.4 Thesis Structure . . . 4
- 1.5 Publications . . . 5

2 Preliminaries **6**
- 2.1 Lattice Theory . . . 6
  - 2.1.1 Sets . . . 6
  - 2.1.2 Tuples . . . 7
  - 2.1.3 Relations . . . 8
  - 2.1.4 Partially Ordered Sets . . . 8
  - 2.1.5 Lattices . . . 9
- 2.2 Formal Concept Analysis . . . 10
- 2.3 The Semantic Web . . . 12
  - 2.3.1 RDF . . . 12
  - 2.3.2 RDF Data Model . . . 13
  - 2.3.3 RDF Vocabularies . . . 13
  - 2.3.4 RDF Schema . . . 15
- 2.4 Characteristic Sets . . . 16
- 2.5 Wikidata . . . 17
- 2.6 Apache Hadoop . . . 19
  - 2.6.1 Hadoop Distributed File System . . . 19
  - 2.6.2 Hadoop MapReduce . . . 19
- 2.7 Linear Regression Models . . . 20
- 2.8 Summary . . . 21

3 Related Work **22**
- 3.1 Computing Formal Concept Lattices . . . 22
- 3.2 Formal Concept Analysis on the Semantic Web . . . 23
- 3.3 Data-Driven RDF schema . . . 24
- 3.4 Modelling dynamics on the Semantic Web . . . 26
- 3.5 Summary . . . 26

4 Proposed Graph Schema **27**
4.1 RDF Concept Lattice . . . 27
4.2 Characteristic Set Lattice . . . 30
4.3 Characteristic Set #-Lattice . . . 31

5 Modelling Dynamics in RDF Graphs **33**
5.1 Characteristic Set Lattice Diff . . . 33
5.2 Adding a Diff to a Lattice . . . 35

6 Extraction of Characteristic Sets **38**
6.1 Architecture . . . 38
6.2 Data . . . 38
6.3 MapReduce Jobs . . . 39
6.4 Optimization and Performance Analysis . . . 40
6.4.1 Number of Reducers . . . 40
6.4.2 OIDs representation . . . 41
6.4.3 Compression . . . 42
6.4.4 Combiners . . . 43
6.4.5 Summary . . . 43
6.5 Characteristic Set Statistics . . . 44

7 Characteristic Set Lattice Computation **46**

8 High Level Dynamics Evaluation **49**
8.1 Flat Prediction . . . 50
8.2 Transitive Prediction . . . 50

9 Conclusions **52**

Bibliography **53**