Abstract: The goal of the Semantic Web is to make the information available on the web easier accessible by providing machine readable meta-data. The Resource Description Framework (RDF) represents the data model proposed for this purpose, and SPARQL the corresponding query language. Despite the need for storing, querying and processing this data efficiently, the progress made towards a deeper understanding of the foundational properties of these technologies can only be considered as first steps towards this goal. Pursuing this work by exploring fundamental problems related to SPARQL query optimization and the efficient storage of RDF data is the main goal of this thesis.

Towards SPARQL query optimization, concentrating on SPARQL’s optionality feature, we initialize the study of static query analysis for the class of well-designed SPARQL queries, and study the complexity of the evaluation, enumeration, and counting problem for those queries. Finally, we present several equivalence preserving transformation rules for this class of queries.

Concerning the efficient storage of RDF data, we study the problem of detecting redundancy in RDF data caused by rule-based inference. We examine a setting including rules and constraints, and also consider redundancy with respect to a set of queries. The result is a fine grained complexity analysis of how the different problem parameters influence the complexity of this problem.

While a central goal of this thesis is to extend results on relational data to RDF and SPARQL, we encounter several interesting problems unanswered even in the relational case. We therefore study the complexity of the model checking problem for several classes of tuple generating dependencies, and initialize the systematic search of tractable fragments of the counting problem for CQs by studying this problem for acyclic CQs.