**BUCHAREST UNIVERSITY OF ECONOMIC STUDIES**

****

Doctoral School Economy

**DOCTORAL THESIS**

**IT SOLUTIONS FOR THE ANALYSIS OF LARGE VOLUMES OF DATA: THE SEMANTIC WEB OF THINGS PARADIGM**

Presented and publicity defended by the author:

GEORGIANA I. STĂNESCU

Scientific supervisor: Prof.univ.dr. ADELA BÂRA

**Bucharest, 2024**

# SUMMARY

The PhD thesis entitled Informatics Solutions for Big Data Analytics: The Semantic Web of Things Paradigm addresses a topical topic, exploring the fundamental concepts of the Semantic Web, its potential benefits, associated challenges, and current state of development. The main objective is to provide solutions to the problems encountered in the Big Data analysis and integration process, capitalizing on the positive effects of the application of semantic technologies.

Given the significant advances in the field of large linguistic models (LLM), this research aims to explore their potential in improving semantic understanding and data retrieval, thus addressing the widespread need for the use of semantic technologies by the majority of users. Although there are numerous articles addressing both semantic technologies (ontologies, semantic markup languages) and Big Data in the same context, their role and application in the web domain has not been specifically highlighted.

During the research, the main semantic technologies and the current state of the use of large volumes of data were studied, the results being presented in articles at international conferences. The thesis is supplemented with case studies that demonstrate the benefits of the Semantic Web in various fields such as education.

The PhD thesis highlights some of the problems associated with Big Data and confirms the need to use semantic technologies for its development and access, by centralizing information in one place. From the topics covered, it can be seen how the demands of society have led to technological developments that contribute to the creation of a much more accessible and usable web environment.

The author's contribution to the research topic consists in proposing IT solutions for the efficient integration and analysis of information in Big Data. The thesis presents a diverse range of data processing techniques, including data collection, cleaning, transformation, integration, storage and analysis, with the aim of facilitating the discovery of new information and supporting decision making.

**TABLE OF CONTENTS**

SUMMARY 2

ACKNOWLEDGMENTS 3

INTRODUCTION 7

CHAPTER 1. THE ISSUE OF THE SEMANTIC WEB: DEFINITIONS, EVOLUTION AND CURRENT STATE 10

1.1 THE SEMANTIC WEB CONCEPT 11

1.1.1 Definition and evolution 12

1.1.2 General aspects of semantic technologies 13

1.1.3 The architecture of the semantic web 15

1.1.4 The impact of the semantic web in various fields 21

1.1.5 Languages ​​specific to semantic technologies 23

1.2 STUDIES AND ANALYSIS CARRIED OUT TO UNDERSTAND THE USE OF SEMANTIC TECHNOLOGIES 30

1.2.1 Trends in the field of semantic technologies 31

1.2.2 A comparative analysis of semantic technologies: advantages and disadvantages 34

1.2.3 Studies and analyzes carried out to understand the use of semantic technologies 38

1.3 ONTOLOGIES – DEFINITION, STRUCTURE, EDITORS, USE 45

1.3.1 General notions about ontologies 45

1.3.2 Editors for creating ontologies 48

1.3.3 Comparison of Ontology Editing Tools 51

1.3.4 Presentation of the e-learning domain modeled with the help of ontologies 52

1.3.5 Using ontologies for semantic data integration 55

CHAPTER 2. ANALYSIS OF LARGE VOLUMES OF DATA 60

2.1 PRESENTATION OF THE CONCEPT OF BIG DATA 60

2.2 STAGE OF RESEARCH IN THE FIELD OF BIG DATA 63

2.3 SEMANTIC TECHNOLOGIES IN THE CONTEXT OF BIG DATA 65

2.3.1 Perspective on the interaction between semantic technologies and big data: current status, impact and research directions 66

2.3.2 The value added by semantic technologies in the context of Big Data 75

2.3.3 Characteristics of semantic databases 82

2.3.4 Study on the use of Big data integration with semantic technologies: data management in the field of e-learning 86

CHAPTER 3. METHODOLOGY PROPOSAL FOR INTEGRATION OF SEMANTIC TECHNOLOGIES IN BIG DATA 96

3.1 STUDIES CARRIED OUT TO UNDERSTAND THE METHODOLOGY AND ALGORITHMS OF INTEGRATION OF SEMANTIC TECHNOLOGIES 96

3.2 ARGUMENTATION OF THE NEED FOR METHODOLOGY FOR THE INTEGRATION OF SEMANTIC TECHNOLOGIES IN BIG DATA 105

3.2 PROPOSED METHODOLOGY: STRUCTURE, ALGORITHM, FUNCTIONALITY 107

3.4.1 Definition of objectives 110

3.4.2 Data collection 112

3.4.3 Choice of technologies 114

3.4.4 Data Analysis 117

3.4.5 Integration and configuration 120

3.4.6 Testing and Validation 124

3.4.7 Monitoring and updating 126

3.4.8 Documentation and optimization 128

3.5 OPTIMIZING THE METHODOLOGY WITH A MATHEMATICAL REPRESENTATION 131

CHAPTER 4. CONTRIBUTIONS TO THE ANALYSIS OF LARGE VOLUMES OF DATA AND THE INTEGRATION OF SEMANTIC TECHNOLOGIES IN BIG DATA 141

4.1 PROPOSED ALGORITHM FOR METHODOLOGY VALIDATION 141

4.2 CASE STUDY: UNIVERSITY WEB PAGES 145

4.2.1 Validation of the proposed methodology 145

4.2.2 The conclusions obtained following the implementation of the proposed methodology 155

CHAPTER 5. RESEARCH CONCLUSIONS AND OWN CONTRIBUTIONS 157

5.1 PRESENTATION OF THE ORIGINAL CONTRIBUTION OF THE WORK 157

5.2 PROSPECTS FOR CONTINUING RESEARCH 159

CHAPTER 6. LIST OF PUBLISHED WORKS 162

7. BIBLIOGRAPHY 164

LIST OF FIGURES 180

LIST OF TABLES 181

LIST OF ABBREVIATIONS 183

**Key words:**

Semantic technologies, Big Data, e-learning, methodology, semantic analysis, databases, ontologies, ontological editors, semantic web, web 3.0