**ACADEMIA DE STUDII ECONOMICE DIN BUCUREŞTI**

****

Şcoala doctorală de Informatică Economică

**TEZĂ DE DOCTORAT**

**SOLUȚII INFORMATICE PENTRU ANALIZA VOLUMELOR MARI DE DATE: PARADIGMA SEMANTIC WEB OF THINGS**

Prezentată şi susţinută public de către autor:

GEORGIANA I. STĂNESCU

Conducător ştiinţific: Prof.univ.dr. ADELA BÂRA

**Bucureşti, 2024**

# REZUMAT

Teza de doctorat intitulată *Soluții informatice pentru analiza volumelor mari de date: paradigma Semantic Web of Things* abordează un subiect de actualitate, explorând conceptele fundamentale ale web-ului semantic, beneficiile sale potențiale, provocările asociate și stadiul actual de dezvoltare. Obiectivul principal este de a oferi soluții la problemele întâmpinate în procesul de analiză și integrare Big Data, valorificând efectele pozitive ale aplicării tehnologiilor semantice.

Având în vedere progresele semnificative în domeniul modelelor lingvistice mari (LLM), această cercetare își propune să exploreze potențialul acestora în îmbunătățirea înțelegerii semantice și în procesul de căutare a datelor, adresând astfel nevoie extinsă de utilizare a tehnologiilor semantice de către majoritatea utilizatorilor. Deși există numeroase articole care abordează atât tehnologiile semantice (ontologii, limbaje de marcare semantică), cât și Big Data în același context, nu a fost evidențiat în mod specific rolul și aplicarea acestora în domeniul web.

Pe parcursul cercetării, au fost studiate principalele tehnologii semantice și stadiul actual al utilizării volumelor mari de date, rezultatele fiind prezentate în articole la conferințe internaționale. Teza este completată cu studii de caz care demonstrează beneficiile web-ului semantic în diverse domenii, cum ar fi educația.

Teza de doctorat subliniază câteva dintre problemele asociate Big Data și confirmă necesitatea utilizării tehnologiilor semantice pentru dezvoltarea și accesarea acestora, prin centralizarea informațiilor într-un singur loc. Din subiectele tratate se observă cum cerințele societății au dus la dezvoltări tehnologice care contribuie la crearea unui mediu web mult mai accesibil și utilizabil.

Contribuția autorului la tema de cercetare constă în propunerea unor soluții informatice pentru integrarea și analizarea eficientă a informațiilor în Big Data. Teza prezintă o gamă variată de tehnici de procesare a datelor, inclusiv colectarea, curățarea, transformarea, integrarea, stocarea și analiza datelor, cu scopul de a facilita descoperirea de informații noi și de a susține luarea deciziilor.

**CUPRINS**

REZUMAT 2

MULȚUMIRI 3

INTRODUCERE 7

CAPITOLUL 1. PROBLEMATICA WEB-ULUI SEMANTIC: DEFINIȚII, EVOLUȚIE ȘI STADIUL ACTUAL 10

1.1 CONCEPTUL DE WEB SEMANTIC 11

1.1.1 Definiție și evoluție 12

1.1.2 Aspecte generale ale tehnologiilor semantice 13

1.1.3 Arhitectura web-ului semantic 15

1.1.4 Impactul web-ului semantic in diverse domenii 21

1.1.5 Limbaje specifice tehnologiilor semantice 23

1.2 STUDII ȘI ANALIZE REALIZATE PENTRU A ÎNȚELEGE UTILIZAREA TEHNOLOGIILOR SEMANTICE 30

1.2.1 Tendințe în domeniul tehnologiilor semantice 31

1.2.2 O analiza comparativa a tehnologiilor semantice: avantaje si dezavantaje 34

1.2.3 Studii si analize realizate pentru a intelege utilizare tehnologiilor semantice 38

1.3 ONTOLOGIILE – DEFINIȚIE, STRUCTURĂ, EDITOARE, UTILIZARE 45

1.3.1 Noțiuni generale despre ontologii 45

1.3.2 Editoare pentru crearea de ontologii 48

1.3.3 Comparație între instrumentele de editare ontologică 51

1.3.4 Prezentarea domeniul e-learning modelat cu ajutorul ontologiilor 52

1.3.5 Utilizarea ontologiilor pentru integrarea datelor semantice 55

CAPITOLUL 2. ANALIZA VOLUMELOR MARI DE DATE 60

2.1 PREZENTAREA CONCEPTULUI DE BIG DATA 60

2.2 STADIUL CERCETĂRILOR ÎN DOMENIUL BIG DATA 63

2.3 TEHNOLOGIILE SEMANTICE ÎN CONTEXTUL BIG DATA 65

2.3.1 Perspectiva asupra interactiunii dintre tehnologiile semantice si big data: stadiul actual, impact si directii de cercetare 66

2.3.2 Valoarea adaugata de tehnologiile semantice in contextul Big Data 75

2.3.3 Caracteristicile bazelor de date semantice 82

2.3.4 Studiu privind utilizare integrarii Big data cu tehnologiile semantice: managementul datelor in domeniul e-learning 86

CAPITOLUL 3. PROPUNEREA METODOLOGIEI PENTRU INTEGRAREA TEHNOLOGIILOR SEMANTICE IN BIG DATA 96

3.1 STUDII REALIZATE PENTRU ÎNȚELEGEREA METODOLOGIEI ȘI ALGORITMILOR DE INTEGRARE A TEHNOLOGIILOR SEMANTICE 96

3.2 ARGUMENTAREA NECESITĂȚII METODOLOGIEI PENTRU INTEGRAREA TEHNOLOGIILOR SEMANTICE ÎN BIG DATA 105

3.2 METODOLOGIA PROPUSĂ: STRUCTURĂ, ALGORITM, FUNCȚIONALITATE 107

3.4.1 Definirea obiectivelor 110

3.4.2 Colectarea datelor 112

3.4.3 Alegerea tehnologiilor 114

3.4.4 Analiza Datelor 117

3.4.5 Integrarea si configurarea 120

3.4.6 Testarea și validarea 124

3.4.7 Monitorizare și actualizare 126

3.4.8 Documentare și optimizare 128

3.5 OPTIMIZAREA METODOLOGIEI CU O REPREZENTARE MATEMATICĂ 131

CAPITOLUL 4. CONTRIBUTII LA ANALIZAREA VOLUMELOR MARI DE DATE ȘI INTEGRAREA TEHNOLOGIILOR SEMANTICE ÎN BIG DATA 141

4.1 ALGORITM PROPUS PENTRU VALIDAREA METODOLOGIEI 141

4.2 STUDIU DE CAZ: PAGINILE WEB ALE SITE-ULUI UNIVERSITĂȚII 145

4.2.1 Validarea metodologiei propuse 145

4.2.2 Concluziile obținute în urma implementării metodologiei propuse 155

CAPITOLUL 5. CONCLUZII ALE CERCETĂRII ȘI CONTRIBUȚII PROPRII 157

5.1 PREZENTAREA CONTRIBUȚII ORIGINALE ALE LUCRĂRII 157

5.2 PERSPECTIVE DE CONTINUARE A CERCETĂRILOR 159

CAPITOLUL 6. LISTA LUCRĂRILOR PUBLICATE 162

7. BIBLIOGRAFIE 164

LISTĂ FIGURI 180

LISTĂ TABELE 181

LISTĂ DE ABREVIERI 183

**Cuvinte cheie:**

Tehnologii semantice, Big Data, e-learning, metodologie, analiză semantică, baze de date, ontologii, editoare ontologice, web semantic, web 3.0