

METODE DE VIZUALIZARE ȘI NAVIGARE PENTRU ANALIZA FORMALĂ A CONCEPTELOR POLIADICE REZUMAT

DIANA TROANCĂ

Analiza formală conceptuală (FCA) face parte din domeniul cunoscut sub numele de Conceptual Knowledge Processing, i.e. procesarea conceptuală a datelor. FCA se ocupă cu înțelegerea faptelor și a corelațiilor într-un set de date și totodată încearcă să ofere explicații pentru existența acestor fapte și legături între date. Analiza formală a conceptelor poliadice este o extensie a cazului clasic care, în loc de relații binare, se bazează pe relații de incidență n -are pentru a defini concepte formale, i.e. clustere de date în care toate elementele sunt corelate prin relația de incidență. Teza “Metode de Vizualizare și Navigare pentru Analiza Formală a Conceptelor Poliadice” oferă o imagine de ansamblu asupra metodelor de vizualizare, navigare și explorare pentru contexte formale poliadice definite ca seturi poliadice de date împreună cu o relație de incidență n -ară.

Cuvintele cheie pentru temele abordate în această teză sunt: Analiză Formală Conceptuală Poliadică, Clustere de Date, Navigare Conceptuală, Web Usage Behavior, Pattern de Navigare (Navigational Pattern), Trendsetter (Inițiator de Comportament) și Answer Set Programming.

Teza are următoarea structură:

I Introducere

- 1 Motivație
- 2 Enunțul Problemei
- 3 Tematica Tezei și Contribuțiile Principale
- 4 Structura Tezei

II Preliminarii și Stadiul Actual al Domeniului de Cercetare

5. Analiză Formală Conceptuală
 - 5.1 Istoric
 - 5.2 Analiză Formală Conceptuală Diadică
 - 5.2.1 Considerații teoretice
 - 5.2.2 Tool-uri și algoritmi
 - 5.3 Analiză Formală Conceptuală Triadică
 - 5.3.1 Considerații teoretice

- 5.3.2 Tool-uri si algoritmi
- 5.4 Analiză Formală Conceptuală Poliadică

6 Teoria Complexității

7 Answer Set Programming

III Vizualizare, Navigare și Explorare în Seturi de Date Poliadice

8 Web Usage Mining folosind Analiză Formală Conceptuală

- 8.1 Web Usage Mining și Web Analytics Metrics
- 8.2 Web Usage Mining pentru Sisteme de E-learning
- 8.3 PULSE - o Platformă Pentru Evaluarea Studenților
- 8.4 Analiza Log-urilor din PULSE folosind Analiză Formală Conceptuală
 - 8.4.1 Preprocesarea Datelor și Descoperirea Pattern-urilor
 - 8.4.2 Analiza și Vizualizarea Pattern-urilor folosind **Circos**
- 8.5 Interpretarea Datelor Triadice folosind **Circos**

9 Clarificarea și Reducerea Contextelor Triadice

10 O Paradigma de Navigare Triadică bazată pe Relația de Reachability

- 10.1 Motivație
- 10.2 Prototip pentru Navigare
- 10.3 Relații de Reachability între Triconcepte
- 10.4 Reachability în Tricontexte Compuse
- 10.5 Proprietăți ale Clusterelor de Reachability
- 10.6 Strategia de Explorare

11 O Paradigmă de Navigare n -adică bazată pe Membership Constraints

- 11.1 Motivație
- 11.2 Membership Constraints
 - 11.2.1 Membership Constraints în Analiză Formală Conceptuală Diadică
 - 11.2.2 Membership Constraints în Analiză Formală Conceptuală Triadică
 - 11.2.3 Membership Constraints în Analiză Formală Conceptuală Poliadică
 - 11.2.4 Discuție asupra Proprietății de Proper Satisfiability
- 11.3 Codarea a Membership Constraints în Answer Set Programming
- 11.4 Navigarea în Spații Conceptuale folosind Membership Constraints
- 11.5 Implementarea Tool-urilor de Navigare și Explorare bazate pe Membership Constraints
 - 11.5.1 ASP Navigation Tool
 - 11.5.2 Brute Force Navigation Tool

11.6 Evaluarea și Compararea Strategiilor bazate pe ASP și Brute Force

IV Concluzii și Planuri pentru Activitatea Viitoare

12 Rezultatele obținute

13 Aspecte Nerezolvate și Planuri pentru Activitatea Viitoare

După cum reiese din structura tezei, Partea a II-a conține capitolele principale ale tezei care includ contribuțiile principale. Aceste contribuții constau în metode noi de vizualizare, navigare și explorare în seturi n -adice de date. În cele ce urmează sunt descrise pe scurt capitolele cuprinse în Partea a II-a.

În Capitolul 8 este descrisă o abordare triadică pentru analiza comportamentului studenților pe platforma PULSE. În acest capitol sunt introduse noțiunile de web usage mining și web analytics metrics și, de asemenea, este oferită o explicație a faptului că tehnicile obișnuite de web usage mining nu dau rezultate bune în cazul unui sistem de e-learning. În consecință, propunem analiza formală conceptuală ca tehnică de web usage mining și o folosim într-o analiză detaliată a log-urilor platformei PULSE. În acest sens, sunt descrise în detaliu cele trei etape ale analizei: pre-procesarea datelor, descoperirea pattern-urilor și analiza pattern-urilor. În procesul de descoperire a pattern-urilor se investighează comportamentul studenților în funcție de diferite aspecte temporale ale cursului. În final, aceste pattern-uri sunt vizualizate într-un format circular folosind *Circos*. În interpretarea rezultatelor se identifică trend-setteri și corelații între comportamentul studenților și activitățile din timpul cursului, cum ar fi asignarea proiectelor, examen parțial, examen final.

În Capitolul 9 sunt definite metode de a reduce mărimea unui set de date triadic fără a schimba structura sa conceptuală. În acest scop sunt extinse noțiunile de clarificare și reducere de la contexte diadice la contexte triadice. Totodată este subliniat faptul ca acestea sunt procese importante în etapa de preprocesare a datelor, având în vedere faptul că influențează doar eficiența analizei și nu rezultatele acesteia. În final, se testează procesele de clarificare și reducere pe o bază de date care conține informații despre pacienți bolnavi de cancer. În urma experimentelor putem concluziona faptul că procesele de clarificare și reducere într-un context reduc drastic mărimea unui set de date și ajută la îmbunătățirea performanței a tool-urilor folosite ulterior pentru analiză.

În Capitolul 10 este descrisă prima paradigmă de navigare propusă pentru seturi de date triadice. Capitolul începe cu un exemplu de navigare care motivează metoda aleasă pentru această paradigmă. Aspectele teoretice ale

paradigmei de navigare se bazează pe o relație de adiacență numită reachability, care rezultă din contexte diadice definite corespunzător. Această relație formează clustere de reachability care conțin concepte corelate reciproc prin relația de reachability. Proprietățile relației de reachability și a clusterelor de reachability sunt studiate la nivel teoretic, iar apoi este introdusă strategia de explorare pe baza relației de reachability. Această strategie folosește proiecții diadice pentru a putea folosi avantajele navigării într-o lattice diadică. Ideea navigării este de a oferi o vizualizare locală asupra unei submulțimi de concepte și de a propune metode intuitive de a naviga de la un astfel de view local la altul adiacent. În cele ce urmează sunt analizate proprietățile strategiei de explorare și sunt propuși algoritmi pentru obținerea structurilor necesare pentru a explora spațiul triconceptelor folosind această strategie.

Capitolul 11 conține a doua paradigmă de navigare propusă care are o abordare diferită, și anume restrângerea spațiului de triconcepte în funcție de constrângeri specificate de utilizator. În acest scop, este introdusă problema de satisfiability of membership constraints pentru a putea determina dacă există un concept care include, respectiv exclude, anumite elemente din componentele sale. Totodată se analizează computational complexity pentru aceste probleme, atât pentru cazuri particulare (diadic și triadic), cât și pentru cazul general. Având în vedere că, în general, problema de satisfiability of membership constraints este NP-complete, o codare a acesteia în ASP este propusă cu scopul de a putea folosi tool-uri ASP care sunt optimizate pentru rezolvarea unor astfel de probleme. În cele din urmă, pe baza noțiunii de membership constraints este propusă o nouă paradigmă de navigare care este implementată pentru cazurile diadic, triadic și tetradic. În aceste implementări sunt exploatate diferite strategii bazate fie pe ASP, fie pe o căutare exhaustivă a spațiului de concepte care este obținut în etapa de preprocesare cu ajutorul unui tool FCA extern. Aceste implementări sunt evaluate și comparate, iar totodată sunt prezentate posibile optimizări, generalizări dar și limitări ale fiecărei abordări.

În concluziile tezei este evidențiată importanța contribuțiilor prezentate și sunt sugerate posibile direcții viitoare de cercetare.