

**CONCEPTUAL VISUALIZATION AND NAVIGATION
METHODS FOR POLYADIC FORMAL CONCEPT
ANALYSIS
EXTENDED ABSTRACT**

DIANA TROANĂ

Formal concept analysis (FCA) is the core of Conceptual Knowledge Processing, being closely related to a deeper understanding of existing facts and relationships, while at the same time trying to find explanations for their existence. Polyadic formal concept analysis is an extension of classical FCA that instead of binary relations uses an n -ary incidence relation to define formal concepts, i.e. data clusters in which all elements are interrelated. The thesis “Conceptual Visualization and Navigation Methods for Polyadic Formal Concept Analysis” gives a deeper insight into visualization, navigation and exploration methods for polyadic formal contexts, i.e. polyadic datasets with an n -adic relation.

The key words of the topics presented in the thesis are: Polyadic Formal Concept Analysis, Data Clusters, Web Usage Behavior, Navigational Pattern, Trend-setter, Conceptual Navigation and Answer Set Programming.

The thesis has the following structure:

I Introduction

- 1 Motivation
- 2 Problem Statement
- 3 Thesis Focus and Key Contributions
- 4 Thesis Outline

II Preliminaries and State of the Art

5. Formal Concept Analysis
 - 5.1 History of Formal Concept Analysis
 - 5.2 Dyadic Formal Concept Analysis
 - 5.2.1 Theoretical Foundations
 - 5.2.2 Existing Tools and Algorithms
 - 5.3 Triadic Formal Concept Analysis
 - 5.3.1 Theoretical Foundations
 - 5.3.2 Existing Tools and Algorithms
 - 5.4 Polyadic Formal Concept Analysis

6 Complexity Theory

7 Answer Set Programming

III Visualization, Navigation and Exploration in Polyadic Datasets

8 Formal Concept Analysis for Web Usage Mining

8.1 Web Usage Mining and Web Analytics Metrics

8.2 Web Usage Mining for E-learning Systems

8.3 PULSE - a PHP Utility used in Laboratories for Student Evaluation

8.4 Applying Formal Concept Analysis on PULSE Usage Data

8.4.1 Data Preprocessing and Pattern Discovery

8.4.2 Pattern Analysis and Visualization using **Circos**

8.5 **Circos** Interpretations of Triadic Data

9 Clarification and Reduction of Triadic Contexts

10 A Triadic Navigation Paradigm based on Reachability Relations

10.1 Motivation

10.2 Proof of Concept

10.3 Reachability Relations among Triconcepts

10.4 Reachability in Composed Tricontexts

10.5 Properties of Reachability Clusters

10.6 Exploration Strategy

11 An n -adic Navigation Paradigm based on Membership Constraints

11.1 Motivation

11.2 Membership Constraints

11.2.1 Membership Constraints in Dyadic Formal Concept Analysis

11.2.2 Membership Constraints in Triadic Formal Concept Analysis

11.2.3 Membership Constraints in n -adic Formal Concept Analysis

11.2.4 A Discussion on Proper Satisfiability

11.3 Encoding for Membership Constraints in Answer Set Programming

11.4 Navigation in Conceptual Spaces based on Membership Constraints

11.5 Implementation of Exploration and Navigation Tool based on Membership Constraints

11.5.1 ASP Navigation Tool

11.5.2 Brute Force Navigation Tool

11.6 Evaluation and Comparison of the ASP and the Brute Force Approach

IV Conclusions and Future Work

12 Achievements

13 Open Issues and Future Work

As we can see from the structure, Part II contains the main chapters of the thesis including the key contributions. These contributions consist of new methods of visualization, navigation and exploration in n -adic datasets. In what follows we briefly describe what each chapter from Part II comprises.

In Chapter 8 we describe a triadic approach for studying the web usage behavior of students using an e-learning platform called PULSE. In this chapter we introduce web usage mining and web analytics metrics and furthermore, we explain why most web usage mining techniques do not work as expected on e-learning systems. As a consequence we propose formal concept analysis as a web usage mining technique and use it in order to run a detailed analysis on the logs of the portal PULSE. We offer a detailed description of the three phases of the analysis, mainly: preprocessing, pattern discovery and pattern analysis. During pattern analysis, we consider temporal aspects while investigating the users' behavior. Finally, we visualize the obtained temporal patterns in a circular layout using a tool called *Circos*. Interpreting these results enables us to identify trend-setters among the students and to correlate their behavior to different course-related activities such as exams or tasks' deadlines.

In Chapter 9 we define methods to reduce the size of a triadic dataset without altering its underlying structure. For this purpose we extend the notions of clarification and reduction from the dyadic to the triadic setting and show that these processes are an important part of the preprocessing phase, considering the fact that they have an influence solely on the efficiency and not on the results of any further analysis. Finally, we run experiments on a cancer registry database and conclude herefrom that clarification and reduction can drastically reduce the size of a dataset and hence increase the performance of the tools used for pattern discovery.

Chapter 10 describes the first navigation paradigm that we propose for triadic datasets. We start by presenting a proof of concept on a small dataset in order to motivate the chosen method of navigation. The theoretical aspects of the navigation paradigm focus on a neighborhood notion, called reachability relation, arising from appropriately defined dyadic concept lattices. This relation gives rise to so-called reachability clusters containing mutually reachable concepts. We study properties arising from these notions and describe an exploration strategy based on the reachability relation, which uses conveniently chosen dyadic projections in order to take advantage of the navigation strategies of a dyadic context. The intuition behind is to locally display a smaller part of the space of triconcepts, instead of displaying all of them at once, and then propose an intuitive navigation strategy that allows for moving from one such local view to other adjacent ones. In addition

to formally analyzing the properties of this strategy, we suggest algorithms for producing the structures necessary for browsing the space of triconcepts using theoretically well-understood methods.

Chapter 11 contains a second navigation paradigm based on a different approach. This approach focuses on narrowing down the space of concepts based on constraints specified by the user. For this purpose, we consider the problem of satisfiability of membership constraints in order to determine if a formal concept exists whose components include, respectively exclude, certain elements. We analyze the computational complexity of this problem for particular cases as well as for the general n -adic problem and, given that the satisfiability problem proves to be NP-complete in general, we present an encoding for membership constraints in answer set programming, which will enable us to use highly optimized ASP tools in the implementation. Next we propose a navigation paradigm based on membership constraints and implement it for the dyadic, triadic and tetradic case using different strategies, one based on the proposed ASP encoding and one using an exhaustive search of the whole concept set, precomputed with an external FCA tool. We evaluate and compare the implementations and discuss possible optimizations and generalizations, but also the limitations of each approach.

In the conclusions of the thesis we highlight the importance of the presented contributions and suggest possible directions for future work.