

Investigating Trend-setters in E-learning Systems using Polyadic Formal Concept Analysis and Answer Set Programming*

Sanda Dragoș[†] and Diana Haliță[†] and Diana Troancă[†]

[†]Babeș-Bolyai University, Cluj-Napoca, România

{sanda,dianat}@cs.ubbcluj.ro, diana.halita@ubbcluj.ro

Abstract

Web-based educational systems offer unique opportunities to study how students learn and based on the analysis of the users' behavior, to develop methods to improve the e-learning system. These opportunities are explored, in the current paper, by blending web usage mining techniques with polyadic formal concept analysis and answer set programming. In this research, we consider the problem of investigating browsing behavior by analyzing users' behavioral patterns on a locally developed e-learning platform, called PULSE. Moreover, we investigate users' behavior by using similarity measures of various chains of accessed pages in a tetradic and a pentadic setting. Furthermore, we present in this paper an approach for detecting repetitive behavioral patterns in order to determine trend-setters and followers.

1 Introduction

Nowadays, the educational system consists of two parts: the traditional educational system and the online educational system. Lately, online educational systems show a rapid development, mainly due to the growth of the Internet [Romero and Ventura, 2007]. Analyzing web educational content is extremely important in order to help the educational process.

Online educational systems consist of techniques and methods which provide access to educational programs for students, who are separated by time and space from traditional lectures. These web-based education systems can record the students' activity in web logs, that provide a raw trace of the learners navigation on the site [Romero and Ventura, 2007].

It has been proven that web analytics are not precise enough for the educational content [Macfadyen and Dawson,

*Diana Haliță was supported by a doctoral research made possible by the financial support of the Sectoral Operational Programme for Human Resources Development 2007-2013, co-financed by the European Social Fund, under the project POS-DRU/187/1.5/S/155383 - "Quality, excellence, transnational mobility in doctoral research". Diana Troancă was supported by a one year research grant from DAAD, the German Academic Exchange Service.

2012], as they were designed to be used on e-commerce sites, which have very different structures and requirements. However, web usage mining [Spiliopoulou and Faulstich, 1999] provides important feedback for website optimization, web personalization [Romero *et al.*, 2009] and behavior predictions [Romero *et al.*, 2013].

From the teaching perspective, the online component becomes a natural extension of traditional learning. Therefore, J. Liebowitz and M. Frank define blended learning as a hybrid of traditional and online learning [Liebowitz and Frank, 2010]. There are a variety of blended learning classes in universities. I.-H. Jo *et al.* compare on one hand the case of the discussion-based blended learning course, which involves active learner's participation in online forums, and on the other, the case of the lecture-based blended learning course, which involves submitting tasks or downloading materials as main online activities [Jo *et al.*, 2014]. In their paper, they show that the data collected in the first case can be analyzed in order to predict linear relations between online activities and student performance, i.e. the total score that they obtain. However, in the second case, the same analysis model was not appropriate for prediction.

It has been shown that finding a single algorithm that has the best classification and accuracy for all cases is not possible, even if highly complicated and advanced data-mining techniques are used [Romero *et al.*, 2013]. Thus, offline information such as classroom attendance, punctuality, participation, attention and predisposition were suggested to increase the efficiency of such algorithms.

In our current research, we use formal concept analysis as a technique to discover patterns in the data logs of the educational portal. Formal concept analysis (FCA) is a mathematical theory based on lattices, that is suitable for applications in data analysis [Wille, 1999]. Due to the strength of its knowledge discovery capabilities and the subsequent efficient algorithms, FCA seems to be particularly suitable for analyzing educational sites. For instance, L. Cerulo and D. Distanto research the topic of improving discussion forums using FCA [Cerulo and Distanto, 2013; Distanto *et al.*, 2014], while our own previous contributions are focusing on applying the same technique in order to analyze the user/student behavior [Dragoș *et al.*, 2014; 2015; 2016].

This paper emphasizes how formal concept analysis tools

along with answer set programming can be used for detecting repetitive browsing habits. The purpose of this research is to determine the following characteristics in the data:

- *trend-setters*, i.e., users which firstly adhere to a specific behavior and then generate a bundle of users following them
- *followers*, i.e., users who copy the behavior of a trend-setter
- *patterns* revealed by the occurrences of particular behaviors (in which weeks, for what trend-setter and what followers)

However, in order to determine trend-setters, we need to look at the data from a different perspective than the ones research in our previous work. With that purpose, we analyze our data from a 4-adic and 5-adic perspective.

2 Polyadic Formal Concept Analysis

In this section, we briefly present the basic notions of formal concept analysis. The fundamental structures are a *formal context*, i.e. a data set that contains elements and a relation between them, and *formal concepts*, i.e. clusters of data from the defined context.

FCA was introduced by R. Wille and B. Ganter in the dyadic setting, in the form of objects related to attributes [Ganter and Wille, 1999]. In subsequent work, F. Lehmann and R. Wille extended it to a triadic setting, adding the third dimension represented by conditions [Lehmann and Wille, 1995]. Later, G. Voutsadakis further generalized the dyadic and triadic cases to n -adic data sets, introducing the term *Polyadic Concept Analysis* [Voutsadakis, 2002]. Formally, an n -adic formal context is defined as follows.

Definition 1. Let $n \geq 2$ be a natural number. An n -context is an $(n+1)$ -tuple $\mathbb{K} := (K_1, K_2, \dots, K_n, Y)$, where K_1, K_2, \dots, K_n are data sets and Y is an n -ary relation $Y \subseteq K_1 \times K_2 \times \dots \times K_n$.

Formal concepts are defined as maximal clusters of n -sets, where every element is interrelated with all the others.

Definition 2. The n -concepts of an n -context (K_1, \dots, K_n, Y) are exactly the n -tuples (A_1, \dots, A_n) that satisfy $A_1 \times \dots \times A_n \subseteq Y$ and which are maximal with respect to component-wise set inclusion. (A_1, \dots, A_n) is called a proper n -concept if A_1, \dots, A_n are all non-empty.

Example 1. Finite dyadic contexts are usually represented as cross-tables, rows being labeled with object names, columns with attribute names. Intuitively, a cross in the table on the row labeled g and the column labeled m , means that object g has attribute m .

In the triadic case, there is a ternary relation that relates objects to attributes and conditions. Here, the corresponding triadic context can be thought of as a 3D cuboid, the ternary relation being marked by filled cells. Therefore, triadic contexts can be unfolded into a series of dyadic "slices". In the following example, we consider a triadic context (K_1, K_2, K_3, Y) where the object set K_1 consists of users, the attribute set K_2 contains chains of visited pages while the conditions K_3 are the weeks of the semester when

the chain occurred as a user's navigational pattern. For this small selection we obtain a $2 \times 4 \times 2$ triadic context, the "slices" being labeled by condition names.

| | | | | |
|-----------|---|---|---|---|
| w3 | A | B | C | D |
| $LT - LA$ | × | | | |
| $LT - LE$ | × | × | × | |
| w4 | A | B | C | D |
| $LT - LA$ | × | | | × |
| $LT - LE$ | × | × | | |

Figure 1: Visit behavior: user, chain of pages, timestamp

There are exactly six triconcepts of this context, i.e., maximal 3D cuboids full of incidences:

- $(\{LT - LA, LT - LE\}, \{A\}, \{w3, w4\})$,
- $(\{LT - LE\}, \{A, B, C\}, \{w3\})$,
- $(\{LT - LE\}, \{A, B\}, \{w3, w4\})$,
- $(\{LT - LA\}, \{A, D\}, \{w4\})$,
- $(\emptyset, \{A, B, C, D\}, \{w3, w4\})$ and
- $(\{LT - LA, LT - LE\}, \{A, B, C, D\}, \emptyset)$.

The first four of these triconcepts are proper.

3 Answer Set Programming for FCA

In the current paper we use answer set programming (ASP) in order to compute formal concepts in contexts of different dimensions. ASP is a logic programming language that uses a declarative approach to solve problems [Gebser *et al.*, 2012].

In 2015, we proposed an ASP encoding that could be used to compute formal concepts and, if necessary, also add some additional constraints to the concepts [Rudolph *et al.*, 2015]. We briefly describe the intuition behind this encoding and highlight the fact that it is easily extended to n -adic contexts.

Let $\mathbb{K} = (K_1, \dots, K_n, Y)$ be an n -context. The first step encoded in the ASP program resembles "guessing" a formal concept candidate (A_1, \dots, A_n) , by indicating for each element of the context if it is included in the concept or not. The second step encodes the elimination of all the previously generated candidates, for which at least one tuple is not included in the relation, i.e. $A_1 \times \dots \times A_n \not\subseteq Y$. In the next steps, candidates that violate the maximality condition or have one empty component, need to be eliminated, ensuring that all the candidates remaining are proper formal concepts of \mathbb{K} . Finally, in the last step, the subset of concepts is selected, for which additional given constraints hold.

It follows from the description of the ASP encoding that it can be extended to compute formal concepts for any dimension n . After encoding the problem for a particular n -adic case, we used for running the ASP program, mainly for the grounding and solving of the encoded problem, Clingo from the Potassco collection [Gebser *et al.*, 2011], since it is currently the most prominent solver leading the latest competitions [Calimeri *et al.*, 2016].

Furthermore, in our previous work, we presented a tool, called ASP navigation tool¹, that allows navigation through the concept space of dyadic, triadic and tetradic data sets based on the previously described ASP encoding [Rudolph *et al.*, 2016]². This tool is based on membership constraints, which are encoded in the last step of the problem. For navigating with this tool, one has to choose elements of the data set and select whether they should be included or excluded from the concept. By adding such constraints, the tool ensures that, eventually, one will get to a final state of the navigation, which corresponds to a proper formal concept, i.e. a real data cluster. The implementation of the ASP navigation tool is described in more details in our previous work [Rudolph *et al.*, 2015; 2016].

In our current analysis we extend the ASP encoding to pentadic data sets and compute formal concepts in order to analyze correlations between tetradic clusters of data in a 5-adic setting and hence obtain interesting patterns, such as trend-setters. Moreover, after analyzing pentadic patterns, we use the previously mentioned tool to take a closer look at some of the students that stand out in the obtained results.

4 Web usage mining on PULSE

Educational environments can store a huge amount of potential data from multiple sources, with different formats, and with different granularity levels (from coarse to fine grain), or multiple levels of meaningful hierarchy (keystroke level, answer level, session level, student level, classroom level, and school level) [Romero and Ventura, 2013]. Therefore, an important research direction focuses on developing computational theories and tools to assist humans in extracting useful information from the rapidly growing volumes of data [Romero and Ventura, 2007]. In Web Mining, data can be collected server side or client side, through proxy servers or web servers logs.

The logged data needs to be transformed in the data pre-processing phase into a suitable format, on which particular mining techniques can be used. Data-preprocessing contains the following tasks: data cleaning, user identification, session identification, data transformation and enrichment, data integration and data reduction [Romero and Ventura, 2007]. Data cleaning is one of the major pre-processing tasks, through which irrelevant log entries are removed, such as crawler activity. For the next steps of the pre-processing phase, more data transformations are necessary, such as data discretization and feature selection, in order to perform user and session identification, data integration from different sources and to further analyze the data.

The usage/access data considered for this analysis is collected from the web logs of an e-learning portal called PULSE [Dragos, 2007]. PULSE records the entire activity of its users and, although it has more types of users, we are currently interested only in the students' activities. PULSE also records other individual information about students such

as the academic results, or users' attendances to the laboratories, which in our university system is mandatory.

We will briefly present the entities which are representative for our study:

- The **user** is a student that accesses web files through a browser. Users can be uniquely identified by their login ID (educational content on PULSE can be accessed only after a login phase);
- A **session** is an actual HTTP session;
- A **chain** is defined as a chronologically ordered sequence of visited pages during a session;
- The **timestamp** is the date and time of the access.

On students' chains, we used Cosine similarity measure [Gan *et al.*, 2007] in order to find similar patterns of usage behavior. For each student we determine chains of pages visited during a visit/session and associate them to the corresponding week based on the visit's timestamp. The next step of the analysis is to compare the chains of a user amongst each other, in order to determine students' repetitive behavior. Furthermore, we compare chains of different users in order to identify the influence one user may have over another, and to get relevant information for identifying possible *trend-setters*, as defined in our previous work [Dragoş *et al.*, 2015].

For the experiments presented in this paper, we consider a group of students from the same program, studying the same subject. For this group, containing 23 students, we logged every single file access of every student, for a specific subject, over a period of one academic semester.

The pages of the e-learning platform are grouped by their content into classes. Our interests for this analysis focuses on 10 of these classes which contain pages related to the educational content. These classes are described and denoted as follows: information about the lecture and the laboratories (I), lab assignments (LA), information about the practical examination (PE), laboratory examples (LE), theoretical support for all laboratories (LT), information about the written examination (WE), overview information about the test papers given during lectures (LP), details on all lecture test papers (LPs), overview information about the lectures (L), slides and notes for all lecture (Ls).

Using the Cosine similarity measure, we obtain pairs of students having similar behavior, i.e. similar chains. That behavior occurs for each student in a certain week. Thus, we have two students, a behavior, and the corresponding weeks in which the behavior occurred for each student. Therefore, for each student X, we can construct a tetradic context, containing all students that have similar behavior with X as objects, behaviors as attributes and weeks as conditions and states. Herefrom, for a student X, a tetradic concept (A_1, A_2, A_3, A_4) can be understood as follows: all students in A_1 have, in comparison to student X, similar behaviors to the ones described by the chains in A_2 ; however, this behaviors occur in the weeks $w_2 \in A_4$ for the students in A_1 , while for student X they occur in the weeks $w_1 \in A_3$.

In order to reduce the granularity of our behavior, which, at this point, is a chain, we substitute all chains with binary codes denoting the presence in that chain of the 10 access

¹<https://sourceforge.net/projects/asp-concept-navigation>

²paper available at <http://www.cs.ubbcluj.ro/~dianat/publications/RuTr1.pdf>

classes that we are interested in. Moreover, to reduce redundancies, we consider an additional constraint, mainly that the timestamp when the behavior occurs for student X should be previous (or identical) to the timestamps when it occurs for the other students. We eliminate the tuples for which the constraint does not hold, hence making sure that a common behavior appears in the context corresponding to a single user, mainly the user initiating that behavior, and not the students that “learn” that behavior. We will refer to this student as *trend-setter* and to the others having the behavior in common as *followers*. In this context setting, we are able to determine bundles of users that have similar behavior. A similar detailed analysis of such user bundles, based on different techniques, was published in a previous paper [Dragoş *et al.*, 2016].

5 Identifying trend-setters based on navigational patterns

In the current paper we would like to analyze our data from the 5-adic perspective, in order to determine trend-setters, i.e. students that create a behavior that is assimilated by others, influencing them in the way they use PULSE. Our approach is to extend the tetradic concept by aggregating all previously described 4-adic contexts into one 5-adic context. Therefore, we introduce a new dimension, called *state2*, that corresponds to the set of users. In the pentadic context, *state* becomes *state1*, in order to avoid any confusions. For computing the pentadic concepts of the described context, we use the ASP program described in Section 3.

The timestamp constraint mentioned above determines that all concepts will contain followers as objects and trend-setters as *state2*, for a specific behavior in the attribute set. The condition set contains the occurrence weeks of that behavior for the trend-setter, while the *state1* set contains the occurrence weeks of that behavior for the followers.

For this set of experiments³, we consider for each trend-setter the behaviors containing the maximum number of classes, from the 10 classes we are interested in. We will denote this behavior as *rich behavior*.

In the obtained results, we observe several patterns. The first pattern gives more insight in determining the trend-setters of particular behaviors. We observe that part of the data can be grouped by behavior (attribute), the week when it first occurred (condition) and the user for which it occurred (*state2*). A sample of these results is presented in Table 2. Here, student F can be identified as a trend-setter, if, for a particular behavior (e.g. “Ls-LP-LT-LA”), he/she is the first to have that rich behavior, and the other students have an 80% similar behavior⁴ in the same or the following weeks.

Using these criterias for grouping the data, we mainly obtain groups with different behaviors and their corresponding trend-setters. These trend-setters are often different from each other and initiate the behavior in different weeks. However, there is one particular case that stands out and that can

³More details about the analyzed data and the obtained results can be found at <http://www.cs.ubbcluj.ro/~fca/tests-for-ai4km/>

⁴The similarity was performed on actual session chains as defined in Section 4.

| Ob. | Attr. | Cond. (w1) | St.1 (w2) | St.2 |
|------------------|-------------|---------------|--------------|------|
| F, B, D, H, Q | Ls-LP-LT-LA | 4 | 4 | F |
| D | Ls-LP-LT-LA | 4 | 4, 6, 7 | F |
| ... | Ls-LP-LT-LA | 4 | ... | F |
| X | Ls-LP-LT-LA | 4 | 8 | F |
| S | Ls-LP-LT-LA | 4 | 13 | F |
| ... | Ls-LP-LT-LA | 4 | ... | F |
| G | Ls-LP-LT-LA | 8 | 15 | X |
| S | Ls-LP-LT-LA | 13 | 13 | S |
| L,W | Ls-LP-LT-LA | 13 | 14 | S |
| G | Ls-LP-LT-LA | 13 | 15 | S |

Table 1: Sample of 5-adic concepts grouped by behavior, trend-setter and week when it occurred

be observed in the subset of data represented in Table 2. Here, we can see that the behavior “Ls-LP-LT-LA” has three potential trend-setters, mainly students F, X and S. The behavior occurs for student F in week 4, for student X in week 8 and for student S in week 13. Although they can all be seen as trend-setters for a particular group of students, we deduce that the real trend-setter is student F, since he is the first to have this behavior. Moreover, the other two students are considered to be his followers, as it can also be seen in the 4th and 5th line of Table 2.

Another aspect that is notable for all the results is that most behaviors are focused on the *Lecture* and *Laboratory* access classes and these behaviors are initiated either at the beginning of the semester, i.e. in weeks 3 and 4, or towards the end of the semester, i.e. in weeks 10 and 12.

The second pattern, that we observed, represents behaviors which are assimilated by other students, who then enrich this behavior and become themselves trend-setters for the new enhanced behavior. This pattern is depicted in Table 1. Here, we can see that user F is the trend-setter for behavior “WE-PE-LT-LA”. User Q learns this behavior, adds a new access class to it, “LPs”, and becomes trend-setter for the new behavior. Then, we can observe that user U learns the new behavior and becomes a follower of user Q.

Another interesting aspect that can be observed in Table 2 is, that there can be two trend-setters initiating the same behavior. Here, we can observe that students B and Q are both trend-setters for the same behavior “I-WE-PE-Ls-LPs”.

Next, we focus on behaviors that have no followers and it turns out that these behaviors also contain longer chains of distinct classes than behaviors that have followers. The subset of concepts corresponding to these behaviors is represented in Table 3. The results show that these behaviors reoccur only in the same week of their initiation and for the same user. Hence, there are no actual followers for those rich behaviors. This can also be deduced from the fact that the object sets of the concepts contain only the user that initiated that behavior, while for other behaviors, that have followers, these can be seen in the object set (see Table 2). The longest

| Ob. | Attr. | Cond. (w1) | St.1 (w2) | St.2 |
|-----|-----------------|---------------|--------------|------|
| F | WE-PE-LT-LA | 5 | 5 | F |
| E | WE-PE-LT-LA | 5 | 6 | F |
| M | WE-PE-LT-LA | 5 | 9 | F |
| D | WE-PE-LT-LA | 5 | 12 | F |
| Q | WE-PE-LT-LA | 5 | 13 | F |
| U,W | WE-PE-LT-LA | 5 | 14 | F |
| V | WE-PE-LT-LA | 5 | 17 | F |
| Q | LPs-WE-PE-LT-LA | 13 | 13 | Q |
| U,W | LPs-WE-PE-LT-LA | 13 | 14 | Q |
| V | LPs-WE-PE-LT-LA | 13 | 17 | Q |
| U,W | LPs-WE-PE-LT-LA | 14 | 14 | U |

Table 2: Followers that become trend-setters for enhanced behaviors

| Ob. | Attr. | Cond. (w1) | St.1 (w2) | St.2 |
|---------|----------------|---------------|--------------|------|
| B, Q, D | I-WE-PE-Ls-LPs | 16 | 16 | B, Q |
| D | I-WE-PE-Ls-LPs | 16 | 16, 17 | B, Q |
| D, V | I-WE-PE-Ls-LPs | 16 | 17 | B, Q |
| O, Y | I-WE-PE-Ls-LPs | 16 | 20 | B, Q |

Table 3: Behavior initiated by two trend-setters

chain observed here contains 8 classes out of the 10 that we are interested in, and the average length of the chains in the behaviors from Table 3 is 7.

In what follows, we present some statistics regarding the number of followers that each trend-setter has, the number of weeks in which the behavior occurs and the size of the chain in the corresponding behavior, in terms of number of “interesting” classes. These statistics are represented in Figure 3. As it can be seen, student F had the most followers (16 different students) for a particular behavior containing 4 important classes, behavior that reoccurs in 11 weeks. Furthermore, we observe that there seems to be a directly proportional relation between the number of followers and the number of weeks in

| Ob. | Attr. | Cond. (w1) | St.1 (w2) | St.2 |
|-----|-------------------------|---------------|--------------|------|
| H | L-Ls-LP-WE-LT-PE-LA-I | 4 | 4 | H |
| C | L-Ls-LP-LT-LE-LA | 5 | 5 | C |
| C | LPs-WE-LT-LE-PE-I | 11 | 11 | C |
| D | Ls-LPs-WE-LT-LE-PE-LA-I | 12 | 12 | D |
| I | L-Ls-LP-LPs-WE-LT-PE | 14 | 14 | I |
| M | L-Ls-LPs-WE-LT-PE-LA | 9 | 9 | M |

Table 4: Longest chains of classes that occur in student’s behavior

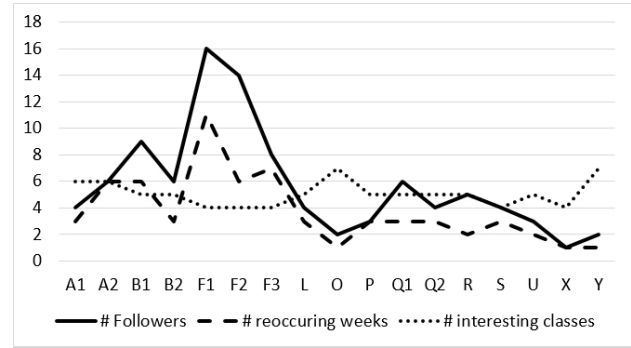


Figure 2: Statistics regarding behaviors and followers

which the behavior reoccurs for each trend-setter. However, there also seems to be an indirect proportional relationship between the number of followers and the size of the chain in a behavior.

In what follows, we focus on user F, which did not stand out in any of the different analyses that we have run on the same data, but using different techniques (presented in our previous work [Dragoş *et al.*, 2015; 2016]). However, this user stands out in the current analysis, for example by having the largest number of followers. The surprise is even greater, since this student attended only 5 out of 14 laboratories and his/her academic results are below average. In order to further analyze the behavior of this particular user, we return to the tetradic approach and use the concept navigation tool based on ASP⁵. Here, we analyze a larger data set as the one used so far, by including all entries that have even small similarities to the behavior of user F, as opposed to a similarity of 80% used in the previous analysis. Next, we navigate to concepts corresponding to the rich behaviors previously observed for user F. Therefore, in the first example, we choose as a first step the behavior “Ls-LP-LT-LA”, which for the data analysis is encoded as “110010010”. Next, we choose the user F as an object, meaning that we are looking for all the weeks when user F repeated this behavior and what other users or behaviors belong to this data cluster. As shown in Figure 5, this behavior occurs only in week 4, but also for users H and D. Moreover, the group of users H, F and D have another behavior in common that occurs in the same week, mainly “LT-LA”, i.e. “10010”. We can see in Figure 5, that although the state represented is an intermediate state, i.e. we did not reach a formal concept yet because the maximality condition is not satisfied, we can already discover patterns in the data.

For the second example, we choose a different rich behavior for F, mainly “LPs-WE-PE-I”, which is encoded as “1100101”, and again student F as an object. This behavior turns out to occur in week 14 and it is a common behavior for users K, H and W. Furthermore, this group of users also has in common the behaviors “PE”, i.e. “100”, and “LPs”, i.e. “1000000”.

Concluding our analysis, we state that trend-setters and followers of particular behaviors can be identified in a pentadic

⁵<https://sourceforge.net/projects/asp-concept-navigation>

| OBJECTS | | ATTRIBUTES | | CONDITIONS | | STATES | | | | | |
|----------------------------------|-----------------------|------------|----------------------------------|-----------------------|-------------|----------------------------------|-----------------------|-----|----------------------------------|-----------------------|-----|
| in | out | in | out | in | out | in | out | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | H | <input checked="" type="radio"/> | <input type="radio"/> | "10010" | <input checked="" type="radio"/> | <input type="radio"/> | "4" | <input checked="" type="radio"/> | <input type="radio"/> | "4" |
| <input type="radio"/> | <input type="radio"/> | Q | <input type="radio"/> | <input type="radio"/> | "110010" | | | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | F | <input checked="" type="radio"/> | <input type="radio"/> | "110010010" | | | | | | |
| <input type="radio"/> | <input type="radio"/> | B | | | | | | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | D | | | | | | | | | |

Figure 3: Generated cluster for behavior “Ls-LP-LT-LA” and user F

| OBJECTS | | ATTRIBUTES | | CONDITIONS | | STATES | | | | | |
|----------------------------------|-----------------------|------------|----------------------------------|-----------------------|-----------|----------------------------------|-----------------------|------|----------------------------------|-----------------------|------|
| in | out | in | out | in | out | in | out | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | K | <input checked="" type="radio"/> | <input type="radio"/> | "100" | <input checked="" type="radio"/> | <input type="radio"/> | "14" | <input checked="" type="radio"/> | <input type="radio"/> | "14" |
| <input checked="" type="radio"/> | <input type="radio"/> | H | <input checked="" type="radio"/> | <input type="radio"/> | "1000000" | | | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | W | <input checked="" type="radio"/> | <input type="radio"/> | "1100101" | | | | | | |
| <input checked="" type="radio"/> | <input type="radio"/> | F | | | | | | | | | |

Figure 4: Generated cluster for behavior “LPs-WE-PE-I” and user F

setting as described earlier in this section. However, in order to take a closer look at the behavior of certain users, it is useful to go back to a tetradic setting and explore correlations of their behaviors and the weeks in which the behaviors occur for the same or for other users. Using the visual navigation tool, one can further explore the data and find potential new patterns which were not revealed by the pentadic context that we have analyzed. Furthermore, the ASP navigation tool can be extended to n -adic datasets, in order to visualize patterns in pentadic or higher-adic contexts.

6 Conclusions and future research

Web is an excellent tool to deliver educational content in the context of an online educational system, while web mining is an efficient technique that can be used to find valuable information in the data. While statistical analysis, through its quantitative approach, might give insight information about web traffic, we believe that formal concept analysis, through its qualitative approach, reveals the potential of hidden patterns inside web logs. Our research is focused on discovering useful patterns that lead to a more efficient interaction between the users and the platform, and that help students acquire the necessary knowledge during the learning process more easily. In this paper, we propose a new method for investigating trend-setters based on pattern extraction from Web log files. We have analyzed students that initiate a behavior that is eventually assimilated by other students, influ-

encing them in the way they use the portal. This analysis helps educators understand the users’ behavior and use the obtained knowledge for optimizing and personalizing the e-learning portal. In our future work, we intend to research whether trend-setters influence the entire evolution of their followers over time. Hence, we need to investigate the evolution of a bundle of users, mainly the followers, over time. In order to deal with the temporal dimension of the data, we plan to apply temporal concept analysis on a conveniently chosen data set.

References

- [Calimeri et al., 2016] Francesco Calimeri, Martin Gebser, Marco Maratea, and Francesco Ricca. Design and results of the fifth answer set programming competition. *Artif. Intell.*, 231:151–181, 2016.
- [Cerulo and Distante, 2013] Luigi Cerulo and Damiano Distante. Topic-driven semi-automatic reorganization of online discussion forums: A case study in an e-learning context. In *Global Engineering Education Conference (EDUCON), 2013 IEEE*, pages 303–310. IEEE, 2013.
- [Distante et al., 2014] Damiano Distante, Alejandro Fernandez, Luigi Cerulo, and Aaron Visaggio. Enhancing online discussion forums with topic-driven content search and assisted posting. In *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pages 161–180. Springer, 2014.
- [Dragoş et al., 2014] Sanda Dragoş, Diana Haliţă, Christian Săcărea, and Diana Troancă. Applying Triadic FCA in Studying Web Usage Behaviors. In *Knowledge Science, Engineering and Management*, pages 73–80. Springer, 2014.
- [Dragoş et al., 2015] Sanda Dragoş, Diana Haliţă, and Christian Săcărea. Attractors in Web Based Educational Systems a Conceptual Knowledge Processing Grounded Approach. In *Knowledge Science, Engineering and Management*, pages 190–195. Springer, 2015.
- [Dragoş et al., 2016] Sanda Dragoş, Diana Haliţă, and Christian Săcărea. Distilling Conceptual Structures from Weblog Data Using Polyadic FCA. In *22nd International Conference on Conceptual Structures, ICCS 2016*, 2016. (in press).
- [Dragos, 2007] Sanda Dragos. PULSE - a PHP Utility used in Laboratories for Student Evaluation. In *International Conference on Informatics Education Europe II (IEEII)*, pages 306–314, Thessaloniki, Greece, November 2007.
- [Gan et al., 2007] Guojun Gan, Chaoqun Ma, and Jianhong Wu. *Data clustering: theory, algorithms, and applications*, volume 20. Siam, 2007.
- [Ganter and Wille, 1999] Bernhard Ganter and Rudolf Wille. *Formal concept analysis - mathematical foundations*. Springer, 1999.
- [Gebser et al., 2011] Martin Gebser, Benjamin Kaufmann, Roland Kaminski, Max Ostrowski, Torsten Schaub, and Marius Thomas Schneider. Potasco: The potsdam answer set solving collection. *AI Commun.*, 24(2):107–124, 2011.
- [Gebser et al., 2012] Martin Gebser, Roland Kaminski, Benjamin Kaufmann, and Torsten Schaub. *Answer Set Solving in Practice*. Synthesis Lectures on Artificial Intelligence and Machine Learning. Morgan and Claypool Publishers, 2012.
- [Jo et al., 2014] Il-Hyun Jo, Yeonjeong Park, Jeonghyun Kim, and Jongwoo Song. Analysis of online behavior and prediction of learning performance in blended learning environments. *Educational Technology International*, 15(2):71–88, 2014.
- [Lehmann and Wille, 1995] Fritz Lehmann and Rudolf Wille. A triadic approach to formal concept analysis. In Gerard Ellis, Robert Levinson, William Rich, and John F. Sowa, editors, *Proceedings of the Third International Conference on Conceptual Structures (ICCS 1995)*, volume 954 of *LNCS*, pages 32–43. Springer, 1995.
- [Liebowitz and Frank, 2010] Jay Liebowitz and Michael Frank. *Knowledge management and e-learning*. CRC press, 2010.
- [Macfadyen and Dawson, 2012] Leah P. Macfadyen and Shane Dawson. Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society*, 15(3):149–163, 2012.
- [Romero and Ventura, 2007] Cristobal Romero and Sebastian Ventura. Educational data mining: A survey from 1995 to 2005. *Expert systems with applications*, 33(1):135–146, 2007.
- [Romero and Ventura, 2013] Cristobal Romero and Sebastian Ventura. Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 3(1):12–27, 2013.
- [Romero et al., 2009] Cristóbal Romero, Sebastián Ventura, Amelia Zafra, and Paul de Bra. Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems. *Computers & Education*, 53(3):828–840, 2009.
- [Romero et al., 2013] Cristobal Romero, Pedro G Espejo, Amelia Zafra, Jose Raul Romero, and Sebastian Ventura. Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education*, 21(1):135–146, 2013.
- [Rudolph et al., 2015] Sebastian Rudolph, Christian Săcărea, and Diana Troancă. Membership constraints in formal concept analysis. In Qiang Yang and Michael Wooldridge, editors, *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*, pages 3186–3192. AAAI Press, 2015.
- [Rudolph et al., 2016] Sebastian Rudolph, Christian Săcărea, and Diana Troancă. Conceptual navigation for polyadic formal concept analysis. In *under review at AI4KM, 4th Workshop on Artificial Intelligence for Knowledge Management, co-located with IJCAI 2016*, 2016.
- [Spiliopoulou and Faulstich, 1999] Myra Spiliopoulou and Lukas C Faulstich. Wum: a tool for web utilization analysis. In *The World Wide Web and Databases*, pages 184–203. Springer, 1999.
- [Voutsadakis, 2002] George Voutsadakis. Polyadic concept analysis. *Order*, 19(3):295–304, 2002.
- [Wille, 1999] R. Wille. *Classification in the Information Age: Proceedings of the 22nd Annual GfKI Conference, Dresden, March 4–6, 1998*, chapter Conceptual Landscapes of Knowledge: A Pragmatic Paradigm for Knowledge Processing, pages 344–356. Springer Berlin Heidelberg, Berlin, Heidelberg, 1999.