



HAL
open science

Highlighting Trend-Setters in Educational Platforms by Means of Formal Concept Analysis and Answer Set Programming

Sanda Dragoş, Diana Şotropa, Diana Troancă

► **To cite this version:**

Sanda Dragoş, Diana Şotropa, Diana Troancă. Highlighting Trend-Setters in Educational Platforms by Means of Formal Concept Analysis and Answer Set Programming. 4th IFIP International Workshop on Artificial Intelligence for Knowledge Management (AI4KM), Jul 2016, New York, NY, United States. pp.71-92, 10.1007/978-3-319-92928-6_5 . hal-01950007

HAL Id: hal-01950007

<https://inria.hal.science/hal-01950007>

Submitted on 10 Dec 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Highlighting Trend-setters in Educational Platforms by Means of Formal Concept Analysis and Answer Set Programming*

Sanda Dragoş, Diana Şotropa and Diana Troancă

Babeş-Bolyai University, Cluj-Napoca, România
sanda@cs.ubbcluj.ro, diana.halita@ubbcluj.ro,
dianat@cs.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. Therefore, we investigate users' behavior by using similarity measures on various sequences of accessed pages in a tetradic and a pentadic setting, and present an approach for detecting repetitive behavioral patterns in order to determine trend-setters and followers. Furthermore, we prove the effectiveness of combining conceptual scale building with temporal concept analysis in order to investigate life-tracks relative to specific behaviors discovered in online educational platforms.

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 [21]. 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 [21].

It has been proven that web analytics are not precise enough for the educational content [19], as they were designed to be used on e-commerce sites, which have very different structures and requirements. However, web usage mining [26] provides important feedback for website optimization, web personalization [23] and behavior predictions [20].

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 [18]. There are a variety of blended learning

* Diana Şotropa was supported by Tora Trading Services private scholarship.

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 [15]. 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 [20]. 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 [28]. 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 [4, 5], while our own previous contributions are focused on applying the same technique in order to analyze the user/student behavior [6–8].

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.

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

Then, we apply methods of Temporal Concept Analysis (TCA) in order to investigate in more detail the behavior of trend-setters and followers throughout the semester. TCA is also used to compare web usage patterns with respect to temporal development and occurrence.

2 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.

2.1 Dyadic Formal Concept Analysis and its Extensions

FCA was introduced by R. Wille and B. Ganter in the dyadic setting, in the form of objects related to attributes [12]. In subsequent work, F. Lehmann and R. Wille extended it to a triadic setting, adding the third dimension represented by conditions [17].

Definition 1. *Dyadic formal context*

A dyadic formal context $\mathbb{K} = (G, M, I)$ is defined as a triple consisting of two sets and a binary relation $I \subset G \times M$ between the two sets. G represents the set of objects, M the set of attributes and I is called the incidence relation. The notation for an element of the incidence relation is gIm or $(g, m) \in I$ and it is read "object g has attribute m ".

In order to define the notion of a formal concept, the derivation operators have to be introduced first.

Definition 2. *The derivation operator*

We define the derivation operator for the object set G and the attribute set M by: $A' = \{m \in M \mid gIm, \forall g \in A\}$ for $A \subseteq G$, and $B' = \{g \in G \mid gIm, \forall m \in B\}$ for $B \subseteq M$. For an element $g \in G$ or $m \in M$, instead of writing $\{g\}'$ and $\{m\}'$, often the notations g' and m' are used.

Based on these derivation operators, the notion of formal concept is introduced.

Definition 3. *Formal concept*

If (G, M, I) is a formal context, then a (dyadic) is defined as a pair (A, B) , with $A \subseteq G$, $B \subseteq M$, $A' = B$ and $B' = A$. A is called extent and B intent of the concept. The set of all concepts of the context (G, M, I) is denoted by $B(G, M, I)$.

Definition 4. *Concept lattice*

Let (G, M, I) be a formal context and (A_1, B_1) , (A_2, B_2) two concepts of this context. (A_1, B_1) is a subconcept of (A_2, B_2) if $A_1 \subseteq A_2$. In this case, (A_2, B_2) is called a superconcept of (A_1, B_1) . The notation is $(A_1, B_1) \leq (A_2, B_2)$. The set of all concepts with this order relation, $(B(G, M, I), \leq)$, is a complete lattice, called the concept lattice of context (G, M, I) .

Definition 5. *Object concept and attribute concept*

Let (G, M, I) be a formal context, $g \in G$ an object and $m \in M$ an attribute. Then, the formal concept (g'', g') is called an object concept and it is denoted by $\gamma(g)$, while the formal concept (m', m'') is called an attribute concept and it is denoted by $\mu(m)$.

Later, G. Voutsadakis further generalized the dyadic and triadic cases to n -adic data sets, introducing the term *Polyadic Concept Analysis* [27]. Formally, an n -adic formal context is defined as follows.

Definition 6. *Polyadic formal context*

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 7. Polyadic formal concept

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	×	×		

Fig. 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.

2.2 Many-valued Contexts

We will briefly recall some definitions introduced by Rudolf Wille in [28] regarding many-valued contexts and conceptual scaling.

Definition 8. Many-valued contexts

A many-valued context (G, M, W, I) consists of sets G , M , and W and a ternary relation I between G , M and W (i.e., $I \subseteq G \times M \times W$) for which it holds that $(g, m, w) \in I$ and $(g, m, v) \in I$ always implies $w = v$.

The triple $(g, m, w) \in I$ is read as "the attribute m has the value w for the object g ". The many-valued attributes can be regarded as partial maps from G in W . Therefore, it seems reasonable to write $m(g) = w$ instead of $(g, m, w) \in I$.

In order to derive the conceptual structure of a many-valued context, we need to scale every many-valued attribute. This process is called *conceptual scaling* and it is always driven by the semantics of the attribute values.

Definition 9. Conceptual scales

A scale for the attribute m of a many-valued context is a formal context $S_m := (G_m, M_m, I_m)$ with $m(G) \subseteq G_m$. The objects of a scale are called **scale values**, the attributes are called **scale attributes**.

Every context can be used as a scale. Formally there is no difference between a scale and a context. However, we will use the term "scale" only for contexts which have a clear conceptual structure and which bear meaning.

The set of scales can then be used to navigate within the conceptual structure of the many-valued context (and the subsequent scaled context). Some scales are predefined (like nominally, ordinally, etc.), while for more complex views, we need to define particular scales.

2.3 Temporal Concept Analysis

Conceptual time systems have been introduced by K. E. Wolff in [29] in order to investigate conceptual structures of data enhanced with a time layer. Basically, conceptual time systems are many-valued contexts, comprising a time part and an event part, which are subject of conceptual scaling, unveiling the temporal development of the analyzed data, object trajectories and life tracks. We briefly recall some basic definitions.

Definition 10. Conceptual Time System

Let G be an arbitrary set, (G, M, W, I_T) and (G, E, V, I_E) many-valued contexts. Let $\{S_m \mid m \in M\}$ be a set of scales for (G, M, W, I_T) , and $\{S_e \mid e \in E\}$ a set of scales for (G, E, V, I_E) . We denote by $T := ((G, M, W, I_T), (S_m \mid m \in M))$ and $C := ((G, E, V, I_E), (S_e \mid e \in E))$ the correspondent scaled many-valued contexts (on the same object set G). The pair (T, C) is called a conceptual time system on G . T is called the time part and C the event part of (T, C) .

Definition 11. Conceptual Time Systems with a Time Relation

Let (T, C) be a conceptual time system on G and $R \subseteq G \times G$. The triple (T, C, R) is called a conceptual time system on G with a time relation.

Definition 12. Transitions in Conceptual Time Systems with a Time Relation

Let (T, C, R) be a conceptual time system on G with a time relation. Then any pair $(g, h) \in R$ is called an R -transition on G . The element g is called the start and h the end of (g, h) .

Definition 13. Conceptual Time Systems with Actual Objects and Time Relation

Let P be a set of objects, G a set of points in time and $\Pi \subseteq P \times G$ a set of actual objects. Let (T, C) be a conceptual time system on Π and $R \subseteq \Pi \times \Pi$. Then the tuple (P, G, Π, T, C, R) is called a conceptual time system on $\Pi \subseteq P \times G$ with actual objects and a time relation R (shortly a CTSOT).

For each object $p \in P$ we can define the set $p^\Pi = \{g \in G \mid (p, g) \in \Pi\}$. Then the set $R_p = \{(g, h) \mid ((p, g), (p, h)) \in R\}$ is called the set of R transitions of p and the relational structure (p^Π, R_p) is called the time structure of p .

Definition 14. Life track of an object

Let (P, G, Π, T, C, R) be a CTSOT and $p \in P$. Then, for any mapping $f: \{p\} \times p^\Pi \rightarrow X$, the set $f = \{(p, g), f(p, g)\} \mid g \in p^\Pi\}$ is called the f -life track of p .

3 Answer Set Programming for FCA

In the current paper we use answer set programming (ASP) as a method of computing formal concepts in contexts of different dimensions. ASP is a logic programming language that uses a declarative approach to solve problems [13].

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 [24]. 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 Clingo from the Potassco collection [14] (since it is currently the most prominent solver leading the latest competitions [3]) for running the ASP program, mainly for the grounding and solving of the encoded problem.

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 [25]. 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

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

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 [24, 25].

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) [22]. 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 [21]. 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 [21]. 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 [10]. 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.

The relationship between different entities can be determined by temporal aspects hidden in the data. Data may describe developments over time or temporal mechanisms

(i.e. time series data) or it may reveal the patterns that evolve over time[16]. Finding evolving patterns is an important challenge which plays a key role in the process of understanding users behavior.

In order to determine users behavioral patterns we consider chains of pages (sequences of visited pages where the accessed page becomes the referrer for the next one). These chains are formed on the assumption that the temporal order of clicks describes the path the user takes through the web site. When the referrer is not the same as the last page accessed it means that the user opened a new browser tab or window, and we called that part of our session chain a new branch.

We have compared chains of the same user in order to determine each student's repetitive behavior. We also compared chains for different users to identify the influence one user may have over another and to get relevant information in order to determine possible trend-setters. For comparing these chains, we have used the Jaccard, Cosine and Sorensen similarity measures [11].

Among these measures, the Cosine Similarity has the advantage of a smaller complexity. Let A and B be two chains, we build the occurrence vectors C_A and C_B for each chain. This similarity measure returns the cosine of the angle between the two occurrence vectors, by the following formula:

$$CosineSimilarity = \frac{C_A \cdot C_B}{\|C_A\| \cdot \|C_B\|}, \quad (1)$$

where $\|C_A\|$ is the magnitude of the vector C_A .

For all these tests, the order in which access classes occurred into a chain has not been taken into consideration. Therefore, having $P_1; P_2$ two pages which were visited by a user in the same session, we consider that "page P_1 is visited just before page P_2 " has the same weight as "the page P_2 is visited just before page P_1 " while we compute chain similarity.

Given that the Cosine Similarity algorithm has the smallest complexity, requires less memory space and it is computational more efficient than the other similarity measures mentioned, we decided to use this method in order to find similar patterns of usage behavior within the same time sequence or among different time sequences. Next we considered only the chains that have a Cosine similarity of at least 80%.

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 [6].

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 in Table 1.

#	Code	Description
1	I	information about the lecture and the laboratories
2	PE	information about the practical examination
3	WE	information about the written examination
4	L	overview information about all the lectures
5	Ls	slides and notes for all lectures
6	LP	overview information about the test papers given during lectures
7	LPs	details on all lecture test papers
8	LA	lab assignments
9	LE	laboratory examples
10	LT	theoretical support for all laboratories

Table 1. Classes of pages we are interested in

The first three classes presented in the Table 1 contain general information related to the way the lecture and laboratories are conducted and about the examination procedure. The next four classes are related to Lectures, while the last three classes are related to the Laboratory activity.

Using the Cosine similarity measure with a threshold of 80%, we obtain pairs of students having similar behavior, i.e. similar chains in a proportion of at least 80%. That behavior occurs for each student in a certain week. Thus, we have pairs of two students, a common 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, the actual 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 $w_1 \in A_3$ for student X, while for the students in A_1 they occur in the weeks $w_2 \in A_4$.

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 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 [7].

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 experiments², 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. In order to analyze each pattern separately, we have grouped the clusters by behavior/pattern (attribute) and sorted them by the week when it first occurred (condition). Thus, we represent each pattern in a different table and observe the corresponding trendsetters. The first such pattern 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 behavior³ 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 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.

² More details about the analyzed data and the obtained results can be found at <http://www.cs.ubbcluj.ro/~fca/tests-for-ai4km/>

³ The similarity was performed on actual session chains as defined in Section 4.

Objects	Attributes	Conditions (w1)	States 1 (w2)	States 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 2. Sample of 5-adic concepts grouped by behavior, trend-setter and week when it occurred

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 3. 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.

Objects	Attributes	Conditions (w1)	States 1 (w2)	States 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 3. Followers that become trend-setters for enhanced behaviors

Another interesting aspect that can be observed in Table 4 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".

Objects	Attributes	Conditions (w1)	States 1 (w2)	States 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 4. Behavior initiated by two trend-setters

Next, we focus on behaviors that have no followers. We call such behaviors *singular*. It turns out that these behaviors contain longer chains of distinct classes than behaviors that have followers, being complex, repetitive behaviors of some students. The subset of concepts corresponding to these behaviors is represented in Table 5. 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 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 5 is 7.

Objects	Attributes	Conditions (w1)	States 1 (w2)	States 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 5. Longest chains of classes that occur in student's behavior

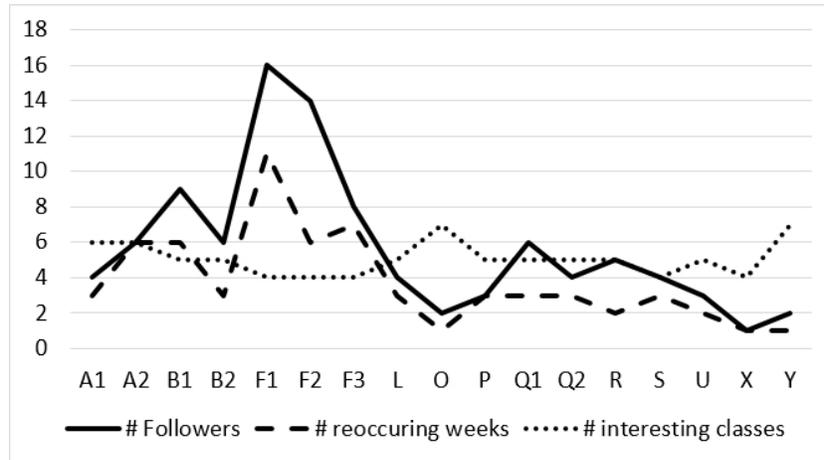


Fig. 2. Statistics regarding behaviors and followers

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 2. Here, for each distinct behavior, i.e., chain, we represented different entries of some users. Therefore, student A has two distinct behaviors on which he/she has followers. We denoted these instances with A1 and A2. Similar, we have two instances for B, three instances for F, and two instances for Q.

As it can be seen in Figure 2, 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 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 [6, 7]). 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 ASP⁴. Therefore, we continue our investigation on the entire data set (i.e., not only the similar entries) in order to determine more details about the behavior of user F. Thus, we navigate through concepts corresponding to the rich behaviors previously observed for user F. As a first example, we start by choosing 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

⁴ <https://sourceforge.net/projects/asp-concept-navigation>

what other users or behaviors belong to this data cluster. As shown in Figure 3, 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 3, that although the state represented is an intermediate state, we can already discover patterns in the data. The fact that it is an intermediate state is determined by the objects which are neither "in", nor "out" of the data cluster, meaning we did not reach a formal concept yet because the maximality condition is not satisfied.

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

Fig. 3. Generated cluster for behavior "Ls-LP-LT-LA" and user F

For the second example, as depicted in Figure 4, 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". Here we have reached a formal concept, as all objects are included in the cluster and there are no more inclusions/exclusions to be determined.

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

Fig. 4. Generated cluster for behavior "LPs-WE-PE-I" and user F

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

Graphically represented conceptual hierarchies prove to be a very efficient tool for the discovery and understanding of complex relationships between knowledge units. ToscanaJ[1, 2] offers the possibility of using available diagrams according to one's interest. Therefore, one can use diagrams aggregation in order to investigate the existence of patterns in attributes correlation. Different scenarios can be formed using only a small subset of the diagrams (the same diagram can even be included more than once).

In another paper ([9]) we have presented an investigation of user behavior in educational platforms using Temporal Concept Analysis, where *attractors* were introduced and defined as sets of scales in conceptual time systems. As the name suggests, an attractor is either influencing or describing the users behavior in the educational platform. Therefore, attractors prove to be special categories of scales which need to be related to specific time granules, when the attractor occurs. Moreover, an attractor represents a specific behavioral pattern. Students, while browsing the e-learning platform, *adhere* to some attractors or not, showing thus particular browsing habits. In this context, we define a behavioral attractor as follows.

Definition 15. Behavioral attractor

A behavioral attractor is a conceptual scale which reflects the habits of a student/user while visiting an e-learning platform at a specific point in time.

Given this definition, the set of all behaviors can be described by a set of conceptual scales on time granules. So, each behavioral pattern represents the event part of the conceptual time system at the specific time granule.

We build users' life tracks by setting time granularity at week level and marking the temporal trajectory of the student through rich behaviors. Then, we consider for every time granule one of the rich behaviors presented in Table 2 and Table 3 (i.e. "Ls-LP-LT-LA", "WE-PE-LT-LA", and "LPs-WE-PE-LT-LA") and build users' life tracks superimposing them on the concept lattice of the corresponding *behavioral attractor*.

In order to emphasize users' life tracks we need to set up some criteria and an order. Then we need to load the appropriate scale in a user defined order. For instance, if we select the following scales in this order: the login scale (a nominal scale which contains user IDs), the weeks scale (a nominal scale which contains the academic weeks), and the corresponding *behavioral attractor*, we obtain a complete view of the trend-setters'

and followers' activity for the three patterns: pattern "Ls-LP-LT-LA" is shown in Figure 5, pattern "WE-PE-LT-LA" in Figure 6, and pattern "LPs-WE-PE-LT-LA" in Figure 7.

Behavioral attractors are *unintended attractors*, which according to our previous research ([9]) are crystallizing behavioral patterns showing how users are using the resources, independently to the intention of the educator. *Behavioral attractor* is an umbrella concept that also encompasses different types of behaviors, some of them defined in our previous paper ([9]), such as *habitual attractor* (i.e., the habit of branching - opening more browser windows or tabs during a visit) and *critical attractor* (i.e., many page accesses that last only a few seconds in critical time intervals such as examination). In this paper we will refer to the navigational (unintended) behavior as to what students are more likely to access during a visit.

In what follows we define a subclass of behavioral attractors called *rich attractors* which expresses the behaviors containing the maximum number of classes.

Definition 16. Rich attractors

A rich attractor is a conceptual scale on a given time granule which reflects the behavior containing the maximum number of classes out of the 10 classes we are interested in (LA, LT, LE, L, Ls, LP, LPs, WE, PE, I). This attractor reveals clues about some other unintended patterns showed off by the users in order to collect educational information.

Further on we use the above described formalizations in order to see in more detail the evolution of the students during the entire semester relative to specific behaviors. Figure 5 presents the first such example. We start with the *rich attractor*, that is the scale corresponding to the behavior ("Ls-LP-LT-LA") generated by the *trend-setter* F. On this scale we build life tracks superimposing the behavior on different weeks on the concept lattice of the rich attractor. Figure 5 contains the life tracks of the *trend-setter* F and two of his *followers*: X and S.

We deduce that the results obtained using TCA and presented in Figure 5 are similar to the ones obtained with the help of Polyadic FCA and presented in Table 2. Figure 5(a) presents the life track of the trend-setter, i.e. how user F is using the educational content considered for this particular rich behavior ("Ls-LP-LT-LA") on the entire semester (i.e., 14 weeks). Analysing the corresponding conceptual scales for users S and X (see Figures 5(b) and 5(c)) which, as we have seen, are followers of user F, gives us insights about the relation between trend-setters and followers. In Figure 5(b) we observe that, after follower S assimilates the trend-setter's behavior in week 13, in week 14 he goes back to his usual behavior and visits only pages from the "LA" class. Follower X, however, adheres to the rich attractor in week 8, and then continues to visit pages from all the classes of the rich behavior "Ls-LP-LT-LA". This indicates that the new behavior of X was, in this case, influenced by this *rich attractor*.

As depicted in Figure 5(b), there are multiple stages in which a user can be placed over time. There are cases in which users are starting by visiting pages from a single class. That is the case of student S which starts visiting one of the classes, i.e. *LT*, included in the considered rich behavior in week 3. This habit is quite normal for a first encounter with the course content. Still, there are situations in which, given a certain point in time, the fact that a student visited pages from a single class might signalize a

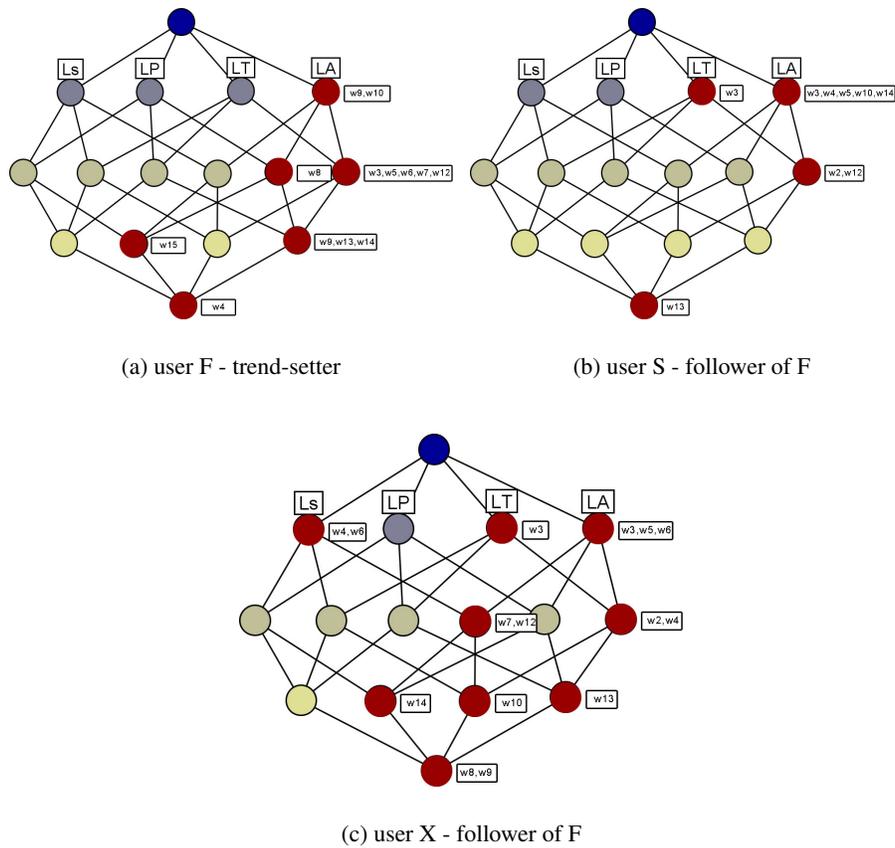


Fig. 5. Users' life tracks according to the sample of 5-adic concepts grouped by behavior and trend-setter as presented in Table 2

superficial approach to the learning process, which needs to be corrected by the instructor. However, the more deeper a life track goes in the concept lattice, the more related content is visited, and thus, it might be assumed that more specific skills related to the subject are acquired and the overview on the learning topic is better. This reflects how seriously users are approaching a specific subject, its structure being unveiled by the corresponding rich attractor. These facts motivate also our interest for rich behaviors.

The *trend-setter* F adheres to the *rich attractor* in the 4th week of the semester. This is represented in the lattice from Figure 5(b) in the lowest node. In the same figure it can be observed that the user F has a tendency to focus more, over time, on lab assignment (LA) pages. He/she maintains over the weeks a more comprehensive behavior than the *followers* (i.e., most of his/her behavior is found on the lower nodes), focusing however on laboratory-related material, and less on the lecture-related classes. The *followers* have a different behavior over the time than the *trend-setter*. However, they seem to have

a very similar behavior among themselves. Moreover, they seem to pay more attention to the lecture-related material.

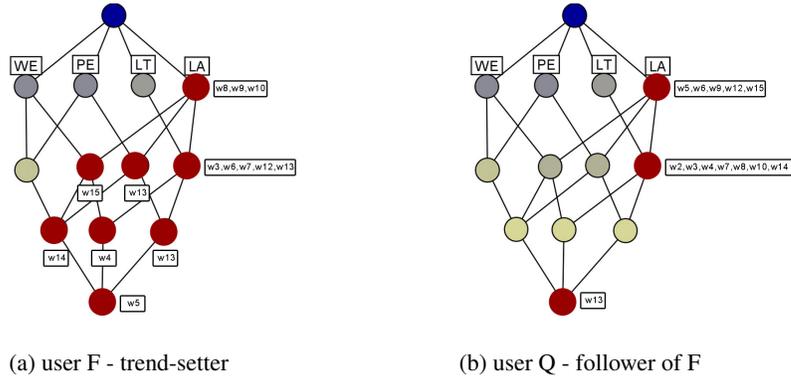


Fig. 6. Users' life tracks according to the sample of 5-adic concepts grouped by behavior and trend-setter as presented in Table 3

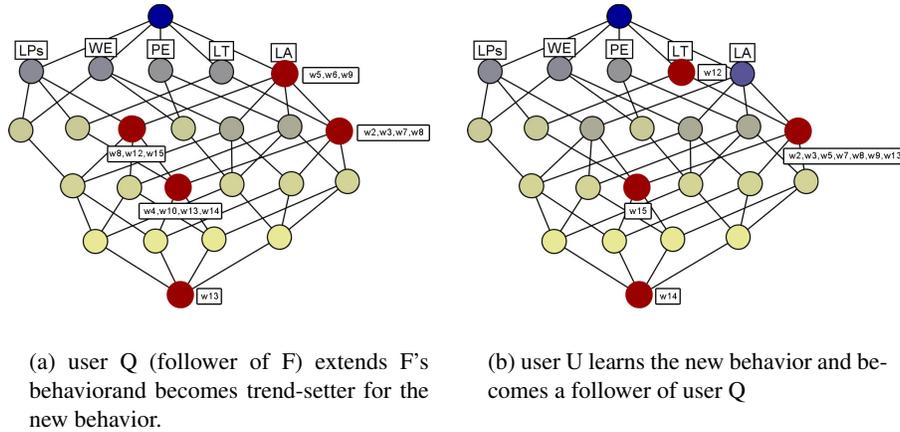


Fig. 7. Users' life tracks of followers that become trend-setters for enhanced behaviors as presented in Table 3

Figure 6 presents the life track of user F as a trend-setter of another rich behavior ("WE-PE-LT-LA") and the life track of user Q, i.e., one of his followers for the corresponding behavior. Moreover, Figure 7 depicts the newly generated rich attractor in which Q becomes the trend-setter of an enhanced behavior ("LPs-WE-PE-LT-LA") and other followers are identified. By looking at Figure 6 one might say that the habits of

the follower seem, at a first glance, not to be influenced by the learned rich behavior. However, by analyzing both figures we can observe how followers learn the behavior of the trend-setter, and then, deviate from the trend-setter's behavior by introducing new rich behaviors.

As depicted in Figure 6(a), *trend-setter* F is interested in all the four classes (i.e., WE, PE, LT, LA) throughout the semester. On the other hand, Figure 6(b) shows that *follower* Q seems to have a “one time” rich behavior in week 13, than returning to his old patterns of accessing only pages related to Laboratories (i.e., LT and LA). However, if we project Q's behavior on the new attractor (as depicted in Figure 7(a)), it can be observed that Q's behavior contains in the latter weeks pages form the Lecture Papers (LPs) class and not only pages related to Laboratories.

User U, follower of Q on the extended behavior "LPs-WE-PE-LT-LA" (see Figure 7(b)) has again a less comprehensive behavior as Q, apart from weeks 14 and 15.

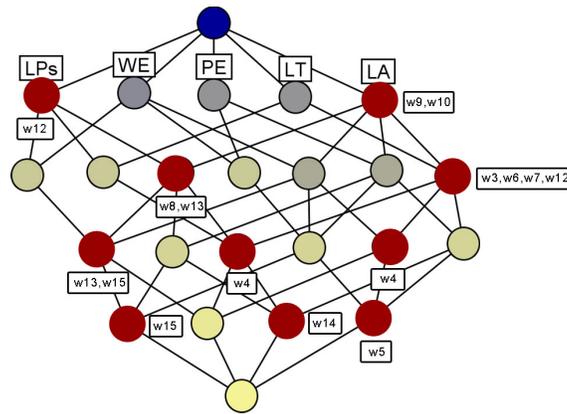


Fig. 8. user F; the initial trend-setter on the extended behavior

If we project the behavior of F on the extended Q's behavior, as depicted in Figure 8, we see that although he/she has visited pages in LPs class, it has no visit containing pages from all 5 classes. However, F has a comprehensive behavior as his behavioral patterns are represented on the lower nodes of the lattice.

7 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, influencing 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. We have also investigated how new navigational patterns initiated by the trend-setters influence the behavior of the followers in time. Moreover, we analyzed the evolution of a bundle of users, over time by applying temporal concept analysis on the data set corresponding to the users that showed up in previous tests. Life tracks give valuable feedback to the instructor regarding how the online educational resources are used over time. They also might be helpful for analyzing the usability of the online educational content, and eventually for improving the structure of the platform and developing new educational instruments. We intend to continue this research, considering pattern structures, relational FCA and other FCA varieties.

References

1. Becker, P., Correia, J.H.: The toscanaj suite for implementing conceptual information systems. In: *Formal Concept Analysis, Foundations and Applications*. pp. 324–348 (2005)
2. Becker, P., Hereth, J., Stumme, G.: ToscanaJ: An Open Source Tool for Qualitative Data Analysis. In: Duquenne, V., Ganter, B., Liquiere, M., Nguifo, E.M., Stumme, G. (eds.) *Proceedings of the Workshop Advances in Formal Concept Analysis for Knowledge Discovery in Databases, FCAKDD 2002*, co-located with the European Conference on Artificial Intelligence, ECAI 2002, Lyon, France. pp. 1–2 (2002)
3. Calimeri, F., Gebser, M., Maratea, M., Ricca, F.: Design and results of the fifth answer set programming competition. *Artif. Intell.* 231, 151–181 (2016)
4. Cerulo, L., Distante, D.: 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*. pp. 303–310. IEEE (2013)
5. Distante, D., Fernandez, A., Cerulo, L., Visaggio, A.: Enhancing online discussion forums with topic-driven content search and assisted posting. In: *Knowledge Discovery, Knowledge Engineering and Knowledge Management*, pp. 161–180. Springer (2014)
6. Dragoş, S., Haliţă, D., Săcărea, C.: Attractors in Web Based Educational Systems a Conceptual Knowledge Processing Grounded Approach. In: *Knowledge Science, Engineering and Management*, pp. 190–195. Springer (2015)
7. Dragoş, S., Haliţă, D., Săcărea, C.: Distilling Conceptual Structures from Weblog Data Using Polyadic FCA. In: *22nd International Conference on Conceptual Structures, ICCS 2016 (2016)*, (in press)
8. Dragoş, S., Haliţă, D., Săcărea, C., Troancă, D.: Applying Triadic FCA in Studying Web Usage Behaviors. In: *Knowledge Science, Engineering and Management*, pp. 73–80. Springer (2014)
9. Dragoş, S., Săcărea, C., Şotropa, D.: An investigation of user behavior in educational platforms using temporal concept analysis. In: *accepted for publication in Proceedings of the 14th International Conference on Formal Concept Analysis (2017)*
10. Dragos, S.: PULSE - a PHP Utility used in Laboratories for Student Evaluation. In: *International Conference on Informatics Education Europe II (IEEII)*. pp. 306–314. Thessaloniki, Greece (2007)

11. Gan, G., Ma, C., Wu, J.: Data clustering: theory, algorithms, and applications, vol. 20. Siam (2007)
12. Ganter, B., Wille, R.: Formal concept analysis - mathematical foundations. Springer (1999)
13. Gebser, M., Kaminski, R., Kaufmann, B., Schaub, T.: Answer Set Solving in Practice. Synthesis Lectures on Artificial Intelligence and Machine Learning, Morgan and Claypool Publishers (2012)
14. Gebser, M., Kaufmann, B., Kaminski, R., Ostrowski, M., Schaub, T., Schneider, M.T.: Potassco: The potsdam answer set solving collection. *AI Commun.* 24(2), 107–124 (2011)
15. Jo, I.H., Park, Y., Kim, J., Song, J.: Analysis of online behavior and prediction of learning performance in blended learning environments. *Educational Technology International* 15(2), 71–88 (2014)
16. Kriegel, H.P., Borgwardt, K.M., Kröger, P., Pryakhin, A., Schubert, M., Zimek, A.: Future trends in data mining. *Data Mining and Knowledge Discovery* 15(1), 87–97 (2007)
17. Lehmann, F., Wille, R.: A triadic approach to formal concept analysis. In: Ellis, G., Levinson, R., Rich, W., Sowa, J.F. (eds.) *Proceedings of the Third International Conference on Conceptual Structures (ICCS 1995)*. LNCS, vol. 954, pp. 32–43. Springer (1995)
18. Liebowitz, J., Frank, M.: Knowledge management and e-learning. CRC press (2010)
19. Macfadyen, L.P., Dawson, S.: Numbers are not enough. why e-learning analytics failed to inform an institutional strategic plan. *Educational Technology & Society* 15(3), 149–163 (2012)
20. Romero, C., Espejo, P.G., Zafra, A., Romero, J.R., Ventura, S.: Web usage mining for predicting final marks of students that use moodle courses. *Computer Applications in Engineering Education* 21(1), 135–146 (2013)
21. Romero, C., Ventura, S.: Educational data mining: A survey from 1995 to 2005. *Expert systems with applications* 33(1), 135–146 (2007)
22. Romero, C., Ventura, S.: Data mining in education. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery* 3(1), 12–27 (2013)
23. Romero, C., Ventura, S., Zafra, A., Bra, P.d.: Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems. *Computers & Education* 53(3), 828–840 (2009)
24. Rudolph, S., Săcărea, C., Troancă, D.: Membership constraints in formal concept analysis. In: Yang, Q., Wooldridge, M. (eds.) *Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, IJCAI 2015, Buenos Aires, Argentina, July 25-31, 2015*. pp. 3186–3192. AAAI Press (2015)
25. Rudolph, S., Săcărea, C., Troancă, D.: Conceptual navigation for polyadic formal concept analysis. In: *Proceedings of the 4th International Workshop on Artificial Intelligence for Knowledge Management AI4KM at IJCAI*. pp. 35–41 (2016)
26. Spiliopoulou, M., Faulstich, L.C.: Wum: a tool for web utilization analysis. In: *The World Wide Web and Databases*, pp. 184–203. Springer (1999)
27. Voutsadakis, G.: Polyadic concept analysis. *Order* 19(3), 295–304 (2002)
28. Wille, R.: Classification in the Information Age: Proceedings of the 22nd Annual GfKI Conference, Dresden, March 4–6, 1998, chap. *Conceptual Landscapes of Knowledge: A Pragmatic Paradigm for Knowledge Processing*, pp. 344–356. Springer Berlin Heidelberg, Berlin, Heidelberg (1999)
29. Wolff, K.E.: Temporal concept analysis. In: *ICCS-2001 International Workshop on Concept Lattices-Based Theory, Methods and Tools for Knowledge Discovery in Databases*, Stanford University, Palo Alto (CA). pp. 91–107 (2001)