STUDIA UNIV. BABEŞ–BOLYAI, INFORMATICA, Volume **LIX**, Number 2, 2014

AN FCA GROUNDED STUDY OF USER DYNAMICS THROUGH LOG EXPLORATION

SANDA DRAGOŞ, DIANA HALIŢĂ, CHRISTIAN SĂCĂREA, AND DIANA TROANCĂ

Abstract. Nowadays the use of e-learning systems is increasing, especially in academia. Most researchers use Web Usage Mining (WUM) techniques to study the behavior of users of such systems. The results thus obtained are used to enhance e-learning systems or make different predictions. Current tools used for Web Usage Mining can offer a partial view on the usage of such systems. This paper presents the use of Formal Concept Analysis (FCA) as a Web Usage Mining instrument. It is also shown how different methods of FCA can be applied for revealing user behavior patterns within a system as well as users dynamics. The detection of users dynamics is a current research topic used to reveal how a system (e-learning instrument in our case) is built. Various authors who use FCA for Web Usage Mining propose ontology generation, semantic web personalisation, but none are using the whole range of FCA tools for exploration. In this paper, we debut by representing user dynamics patterns using ToscanaJ. Then, we use a triadic approach in order to investigate the behavioral patterns of students using the e-learning system. We also show how frequent triconcepts can be used for Web Usage Mining. Then, we use CIRCOS to represent the information content of the triadic data selection.

1. Introduction

Browsing web pages became an essential aspect of our everyday life. The World Wide Web developed to a huge information service center, having an explosive growth of information. Navigation on the Internet has more and more a social dimension ([5]) and becomes a very effective mechanism for acquiring knowledge ([3]). Improving web communication is an essential task, in order to adapt the content and to make it more feasible with the expectations of the visitors, and, on the other hand, to improve both content and design of

Received by the editors: July 3, 2014.
2010 *Mathematics Subject Classification.* 68T30, 68P20.
1998 *CR Categories and Descriptors.* H.3.5 [**Information Systems**]: Information Storage and Retrieval – *Online Information Services*; I.2.6 [**Computing Methodologies**]: Learning – *Knowledge acquisition*.
Key words and phrases. Web Usage Mining, Formal Concept Analysis, Triadic FCA.

the web pages. Behavioral aspects become more and more important, and web usage tools have been developed in order to analyze the web usage behavior. Even if the Internet is nowadays a popular resource, there are many unknown aspects about the intrinsic properties of the Web. Recent investigations try to unveil the way how people interact with the Internet, its impact on our daily life, the cultural transformations which come along with it ([2], [22], [12]).

A special attention has to be paid to educational webpages. These are not built for entertainment, even if at least some sort of entertainment should always be connected with learning. Educational web products, or learning environments must provide a clear structure, their content must be organized and navigation has to be simple and easy to perform. Hence, satisfying the objectives of the web site has become our major concern. Initially emerged from the corporate environment, but then rapidly adapted to any other needs, web analytics and its newer research field, web usage mining, offer many communication analysis techniques ([17]).

A large amount of collateral information about web usage information is stored in databases or web server logs. Statistics and/or data mining techniques are used to discover and extract useful information from these logs ([15]). Other techniques include predictions about the user's behavior by discovering of interesting web usage patterns ([21]).

Web analytics tools are based on some web analytics metrics. They prove to be a proper method to give a rough insight about the analyzed web site, especially if it is a commercial site. Nevertheless, there are situations where these tools are quite inefficient. For instance, the behavior of users in e-learning environments is driven by information acquiring. The learning process takes time, and therefore a visit on an educational site does not apply to the heuristics used by most analytics instruments ([7]). Moreover, as highlighted in [17], higher levels in organizations need summarized and conceptual information for decision making. At these levels, the results produced by web analytics fail this target, since their results target web designers and web developers.

On the other hand, web usage mining [20] focused mainly on research, like pattern analysis, system improvement, business intelligence, usage profiles ([21]). The process of web usage mining undergoes three phases: preprocessing, pattern discovery, and pattern analysis. The preprocessing phase is essential, since at this stage data is cleaned, the users and the sessions are identified and the data is prepared for pattern discovery. Such usage analysis constitutes an important feedback for website optimization, but they are also used for web personalization [19] and predictions [18].

To the best of our knowledge, Formal Concept Analysis was used in Web Usage Mining for semantic web personalization [23, 24].

The same data we are now analyzing (i.e., collected by the usage of PULSE portal) was also analyzed using web analytics [9], web mining [7], visualization using force directed graphs [8] and Formal Concept Analysis [10]. The research presented in this paper is due to the fact that we are still searching for that instrument that offers more insightful findings.

We propose a Conceptual Knowledge Processing grounded approach. The data is preprocessed and all the information which is stored in the database is reorganized according to some conceptual issues. The result is then interpreted with the TOSCANA suite, scales are built and different browsing scenarios are set up. The result is then analyzed and the interesting facts are highlighted.

Besides that, a novel approach is presented, based on a triadic interpretation of the web usage data. Based on the TRIAS algorithm developed by R. Jaeschke [13], we have built a tool called Toscana2TRIAS which connects to a database and enables the user to define the objects, attributes and conditions. Moreover, if a conceptual schema has been built upon the data set, i.e., the data has been preprocessed for ToscanaJ [1], then the user has even more control over the selection of the objects, attributes and conditions. From the conceptual schema, a part of the scaled attributes can be considered as conditions, the rest being considered as attributes in the tricontext. The ternary relation is then automatically read from the database, triconcepts [16] are built and displayed. Since triconcepts contain a large amount of data, displaying their intent, extent and modi is not helpful for a conceptual web usage interpretation. Hence, we have used CIRCOS [4] to display relevant connections in the triadic dataset. To our best knowledge, this is the first attempt to study the user dynamics through log exploration using triadic FCA.

2. Formal Concept Analysis

In the following, we briefly recall some definitions. For more, please refer to [11]. Formal Concept Analysis (FCA) is a mathematical theory which appeared at the end of the 1980's in order to restructure lattice theory in a form that is suitable for applications in data analysis. Formal Concept Analysis is a mathematisation of the traditional understanding of a concept of being a unit of thought, which is homogeneous and coherent.

The fundamental data structure FCA uses is *a formal context*, which exploits the fact that data is quite often represented in two dimensional tables of objects and attributes. Hence, a *formal context* is a triple (G, M, I) consisting of two set, a set of *objects*, G, and a set of attributes, M, and a binary relation $I \subseteq G \times M$, called the *incidence relation*. The fact that gIm is read *object* g *has attribute* m.

Formal contexts can be represented as cross-tables, the rows of which are representing objects, the columns attributes, while the incidence relation is represented by crosses in that table.

We define a Galois-connection (called in FCA the derivation operator) between the power set of G and the power set of M by

$$A' := \{m \in M \mid \forall g \in A.gIm\}, A \subseteq G,$$

and

$$B' := \{g \in G \mid \forall m \in B.gIm\}, B \subseteq M.$$

If $\mathbb{K} := (G, M, I)$ is a formal context, then a formal concept is defined to be a pair (A, B), with $A \subseteq G$ and $B \subseteq M$ and $A' = B, B' = A$. We denote the set of all formal concepts of a formal context $\mathbb{K}$ by $\mathfrak{B}(\mathbb{K})$. This set can be ordered by the *subconcept-superconcept* relationship

$$(A, B) \leq (C, D) :\Leftrightarrow A \subseteq C (\Leftrightarrow D \subseteq B).$$

The subconcept - superconcept relationship $(\leq)$ is an order on $\mathfrak{B}(\mathbb{K})$ and $(\mathfrak{B}(\mathbb{K}), \leq)$ is a complete lattice, called *conceptual hierarchy.* This conceptual hierarchy can be graphically represented by an order diagram, each node representing a concept, while the lines are representing the order relation.

2.1. **Triadic FCA.** Rudolf Wille and Fritz Lehmann extended Formal Concept Analysis in 1995 [16] with Triadic Formal Concept Analysis, considering objects, attributes and conditions:

Definition 1. *A triadic formal context (shortly tricontext) is a quadruple* $\mathbb{K} := (K_1, K_2, K_3, Y)$ *where* K_1, K_2 *and* K_3 *are sets, and* Y *is a ternary relation between them, i. e.,* $Y \subseteq K_1 \times K_2 \times K_3$*. The elements of* K_1, K_2 *and* K_3 *are called (formal) objects, attributes, and conditions, respectively. An element* $(g, m, b) \in Y$ *is read object* g *has attribute* m *under condition* b*.*

As in the dyadic case, derivation operators are defined, in order to determine the conceptual structure of a tricontext, i.e., the triconcepts.

Definition 2. *A triconcept of a tricontext* $\mathbb{K} := (K_1, K_2, K_3, Y)$ *is a triple* (A_1, A_2, A_3) *with* $A_i \subseteq K_i$*, respectively, such that it is maximal with respect to component-wise set inclusion in satisfying* $A_1 \times A_2 \times A_3 \subseteq Y$*.*

If $\mathbb{K} := (K_1, K_2, K_3, Y)$ is described by a three-dimensional cross table, a triconcept is represented by a maximal rectangular box full of crosses, under suitable permutations of rows, columns and layers of the cross table. For a particular triconcept (A_1, A_2, A_3), the components A_1, A_2, and A_3 are called the *extent*, the *intent*, and the *modus* of (A_1, A_2, A_3), respectively.

Definition 3. *For $\{i, j, k\} = \{1, 2, 3\}$ with $j < k$ and for $X \subseteq K_i$ and $Z \subseteq K_j \times K_k$, the $(-)^{(i)}$-derivation operators are defined by*

$$X \mapsto X^{(i)} := \{(a_j, a_k) \in K_j \times K_k \mid (a_i, a_j, a_k) \in Y \textit{ for all } a_i \in X\},$$

$$Z \mapsto Z^{(i)} := \{a_i \in K_i \mid (a_i, a_j, a_k) \in Y \textit{ for all } (a_j, a_k) \in Z\}.$$

These derivation operators correspond to the derivation operators of the dyadic contexts defined by $\mathbb{K}^{(i)} := (K_i, K_j \times K_k, Y^{(i)})$, where

$$a_1 Y^{(1)} (a_2, a_3) \Leftrightarrow a_2 Y^{(2)} (a_1, a_3) \Leftrightarrow a_3 Y^{(3)} (a_1, a_3) \Leftrightarrow (a_1, a_2, a_3) \in Y.$$

Definition 4. *For $\{i, j, k\} = \{1, 2, 3\}$ and for $X_i \subseteq K_i, X_j \subseteq K_j$ and $A_k \subseteq K_k$, the $(-)^{A_k}$-derivation operators are defined by*

$$X_i \mapsto X_i^{A_k} := \{a_j \in K_j \mid (a_i, a_j, a_k) \in Y \textit{ for all } (a_i, a_k) \in X_i \times A_k\},$$

$$X_j \mapsto X_j^{A_k} := \{a_i \in K_i \mid (a_i, a_j, a_k) \in Y \textit{ for all } (a_j, a_k) \in X_j \times A_k\}.$$

These derivation operators correspond to the derivation operators of the dyadic contexts defined by $\mathbb{K}^{ij}_{A_k} := (K_i, K_j, Y^{ij}_{A_k})$ where

$$(a_i, a_j) \in Y^{ij}_{A_k} \Leftrightarrow (a_i, a_j, a_k) \in Y \text{ for all } a_k \in A_k.$$

3. Web Usage Mining using ToscanaJ

Formal Concept Analysis is, by now, an established method for mining knowledge in a variety of fields. We are interested mainly in understanding the conceptual structures of web usage behavior, rather than the use of classical statistical, web analytics or data mining techniques. Hence, using FCA and the knowledge management tool ToscanaJ lies at hand.

The ToscanaJ knowledge management suite ([1]) is specifically tailored for knowledge representation and offers a powerful tool to understand the connections between stored information. The ToscanaJ suite has been applied to a large variety of data sets and used in hundreds of projects. Even if the number of projects involving ToscanaJ has decreased in the past years, due to some flaws in the design of this suite, the efficiency in representing complex knowledge has not been equaled by any other FCA tool developed since then.

The web site used for collecting the usage/access data is an e-learning portal called PULSE [6]. A detailed description of using ToscanaJ to build a conceptual information system for PULSE is given in [10]. Here we focus mainly on those scales which will be later used by Toscana2TRIAS in order to build tricontexts.

The analysis is performed on the data collected from the second semester of the academic year 2012-2013 (i.e., from the beginning of February 2013 to the end of July 2013). A log system records all PULSE accesses into a

MySQL database. For the analized time interval there were 40768 PULSE accesses. The collected information contains the following data fields: full request-URI(including the domain, the requested URL, and any applicable query parameters), referrer URL, cookie ID, login ID, the time stamp of the request, the IP address of the originating web page request, full unmodified User-Agent string, client Screen Resolution.

Using ToscanaJ, a conceptual information system has been built over the PULSE log files. Each data field has been scaled and a conceptual scale has been created. The datasets are considered many-valued contexts, the semantics of attributes being expressed by *conceptual scales*. When building a conceptual scale, Toscana uses SQL statements to gather the required information from the database. Some scales (nominal, ordinal, grid) are already implemented and are created automatically by the software.

For the current investigation we use only the first 3 data fields: Full request-URI, Referrer URL, Cookie ID. We intend to use the rest in the future investigations. The data to be analyzed contains 751 distinct request-URIs (i.e., access files), 471 distinct referrers and 3472 distinct cookie IDs.

3.1. **Prerequisite: Extending the log database.** As a prerequisite, we extended the database by adding two more fields to help in our investigation: Access File Class and Referer Class.

3.1.1. *Access File Classes:* The request-URI represents the address of the accessed webpage along with all query information used for that actual request. Although we value the information contained by this field, the granularity of the accessed web pages is too fine for our intent (for instance, there are 751 distinct access file entries in the database). Therefore, the accessed webpages have been divided into 9 classes. These classes have been nominally scaled (see Figure 1).

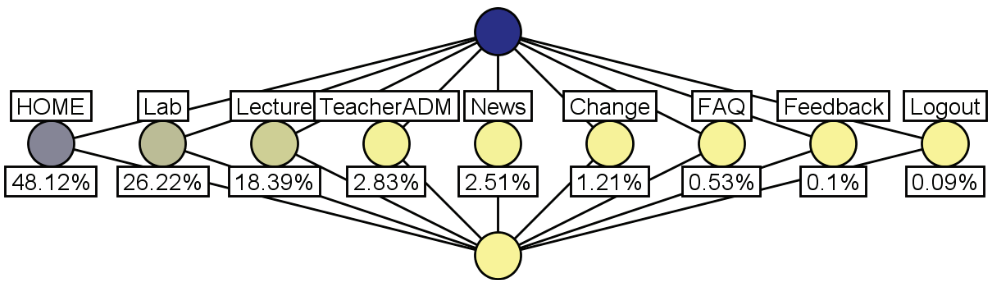


FIGURE 1. The nine Access File classes - a more coarse granularity for access files

PULSE portal was intended to be used mainly during laboratory sessions for students to consult the theoretical support provided by the teacher, to access appointed assignments and check his/her marks and attendances. To distinguish between its users, the portal has a login phase. After login, each user enters the HOME page which contains general information such as: lab attendances, links to laboratory support, links to assignments, marks, evaluation remarks and current announcements. It is expected that the HOME page has the highest percentage of usage being like a "*command center*" of PULSE. The webpages containing the laboratory material come next. Then, there are the web pages containing the material concerning the lectures (e.g., teaching material used for courses, testpaper result, technical solutions).

PULSE provides a set of administrative utilities for the teacher (e.g assign marks, record attendances, post teaching material, add news). All these webpages are grouped into the TeacherADM class.

Along the sections mentioned above, there are also other informative sections for students (i.e news and *frequently asked questions* (FAQ)) as well as the feedback offered by students to a teacher.

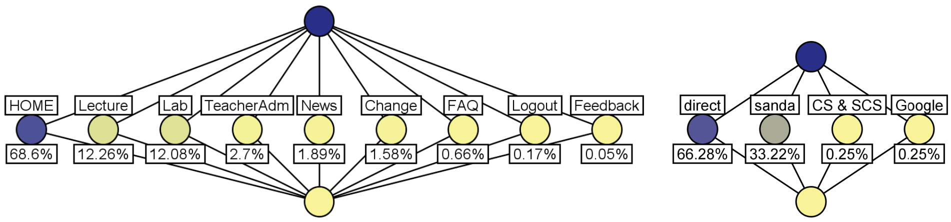
PULSE has a different content for every year of study as the teaching data is updated yearly. Moreover, each year a new set of students would enroll for specific subjects. One student has a single account, however, he/she can navigate through the different subject and/or years of study on which he/she has access on PULSE. The websites from the class CHANGE offer that facility.

The logout is seldom used since the PULSE session can end by simply closing the browser.

3.1.2. *Referrer Classes:* Referrer URLs (on position 2 in the list at the beginning of Section 3) represent the webpage/site from which the current webpage/access file was accessed. As these webpages may also be access files from PULSE or other sites accessed with distinct queries attached to their URL address, the number of these referrers is also high. On the considered database we have 471 distinct referrers. Thus, we also divided these into classes, as described in Table 1.

The referrer classes have also been scaled nominally and visualized by ToscanaJ. These referrer classes can be divided into two disjoint subclasses representing the accesses inside or outside PULSE as depicted in Figure 2. The objects considered were visits identified by cookieIDs.

Inside PULSE, the referrer classes have similar values with their corespondent access file classes. In the subclass of referrers from outside PULSE, the most frequent are direct accesses. That means that students use to save/bookmark the webpages they visit. Therefore it is not advisable to change URI's too often. The link on the teacher website is the second source used.



(a) 95.7% of all referrers are from inside PULSE

(b) 4.3% of all referrers are from outside PULSE

Figure 2. Nominal scales of referrer classes inside/outside PULSE

4. Web Usage Mining with Triadic FCA

The extension Toscana2TRIAS allows the selection of triadic data starting from a given set of scales, if the data has been preprocessed for ToscanaJ. From the conceptual schema file, we have selected the scales presented above, and obtained a triadic data set using Referrer classes (R_class) as attribute set, Access File classes (AF_class) as conditions. For the objects set, we used in turn the database field of Login and IP used by the client originating the web page request. Then, we have generated all triconcepts using TRIAS.

The problem of visualizing triadic data has not been yet satisfactory solved. Triadic conceptual structures have been visualized for instance using trilattices or graphs. CIRCOS as a visualizing tool has been developed to investigate structural patterns arising in bioinformatics. In this section, we present a proof of concepts in order to show possible applications of using CIRCOS to visualize triadic content.

4.1. **Interpreting triadic FCA results with CIRCOS.** CIRCOS is a software package for visualizing data and information in a circular layout. This

Table 1. Referrer classes

Name	Description
inside PULSE	accesses from PULSE webpages
sanda	teacher personal site
cs & scs	faculty site & student site
google	accesses from google search, and/or google mail
direct	direct accesses from bookmarks or by typing the URL of that webpage directly on the browser

circular layout emphasizes patterns in the dataset, showing connections between represented data ([4]).

The input data for CIRCOS is obtained from the tricontext using a derivation operator. We implemented an algorithm that analyzes the XML output of TRIAS and transforms it into a valid input for CIRCOS.

The XML file that results as an output from TRIAS contains all triconcepts which can be derived from the tricontext defined over the data set using Toscana2TRIAS. Each of them is defined by an extent, an intent and a modus. The valid input data set for CIRCOS is a bidimensional table $R \times C$, with numerical values, hence we have to derive these tables from the tricontext. For this, we proceed as follows:

Starting from the tricontext (G, M, B, Y), we first build a dyadic projection $\mathbb{K}_{32} := (G, (B, M), I)$, where $(g, (b, m)) \in I \Leftrightarrow (g, m, b) \in Y$. Then, for each pair (b, m) we compute the corresponding attribute concept $\mu_{\mathbb{K}_{32}}$ and determine the cardinality of its extent $(b, m)'$.

The set of column indicators, denoted C, is the set obtained by projecting the ternary incidence relation Y on M, $\mathrm{pr}_M(Y) := \{m \in M \mid \exists (g, b) \in G \times C.\ (g, m, b) \in Y\}$. Similarly, the set of row indicators, denoted R, is the set obtaining by projecting Y on the set of conditions B.

The algorithm we have implemented builds a table having these sets as column and row indicators, and calculates the numerical values of the table as follows. For each pair $(c, r) \in C \times R$, the cardinality of the extent $(c, r)'$ in $\mathbb{K}_{32}$ is computed directly from the XML output of TRIAS, a small example is given below. This cardinality represents the numerical value of the cell corresponding to the column c and the row r.

TRIAS output example

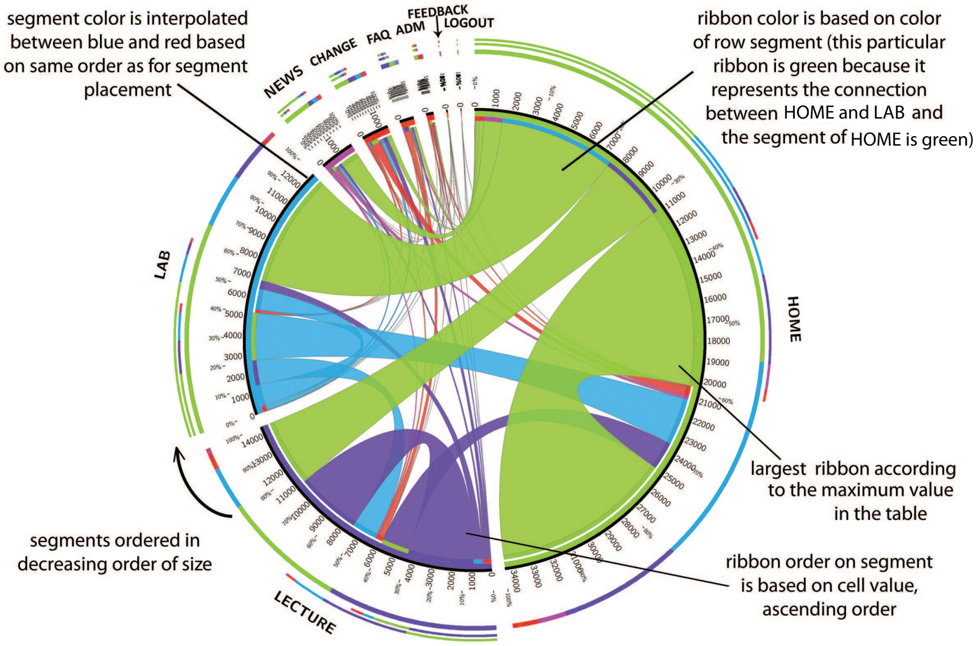
```
<Triconcepts>
  <Triconcept id=”1”>
    <Extent>
      <Object>e1</Object>
      <Object>e2</Object>
    </Extent>
    <Intent>
      <Attribute>i1<Attribute>
      <Attribute>i2<Attribute>
    </Intent>
    <Modus>
      <Condition>m1</Condition>
    </Modus>
  </Triconcept>
```

```
    <Triconcept id="2">
        <Extent>
            <Object>e1</Object>
        </Extent>
        <Intent>
            <Attribute>i2</Attribute>
        </Intent>
        <Modus>
            <Condition>m2</Condition>
            <Condition>m3</Condition>
        </Modus>
    </Triconcept>
</Triconcepts>
```

Corresponding CIRCOS input

```
-   i1  i2
m1  2   2
m2  0   1
m3  0   1
```

As a final step we visualize our data by running CIRCOS and obtaining an output in png or svg format. Figure 3 presents an example for a set of triconcepts having cookie ids as objects, referrer classes as attributes and access file classes as conditions. Because the set of referrer classes and the set of access file classes have elements in common, the sets C and R are not disjoint. As depicted in Figure 3, the segments are represented clockwise starting with the green segment labeled "HOME" at the top of the circle, following with the lila segment labeled "LECTURE", the blue segment labeled "LAB" and the segments "NEWS", "CHANGE", "FAQ", "ADM", "FEEDBACK" and "LOGOUT". The obtained circular layout may be interpreted as described in Figure 3. Each ribbon corresponds to a pair (Referrer class, Access File class). The color of the ribbon is given by the Referrer class segment color. For example the green ribbon from "HOME" to "LECTURE" corresponds to all the pages from Acces File class "LECTURE" accessed from the Referrer class "HOME". For a better understanding we added in Figure 4 a representation in form of a directed graph of the connections between Referrer classes and Access file classes. The nodes of the graph have the same colors as the corresponding segments in the CIRCOS representation. Since Referrer classes and Access file classes are not disjoint, there can be edges directed in both ways between two nodes and also loops.



-	LAB	LECTURE	LOGOUT	FEEDBACK	CHANGE	FAQ	ADM	HOME	NEWS
LAB	1077	1291	0	5	145	59	13	2459	129
LECTURE	450	3420	0	12	231	80	24	1449	122
LOGOUT	0	0	4	1	0	0	0	11	0
FEEDBACK	1	4	0	3	0	0	0	0	3
CHANGE	90	81	0	0	51	0	22	364	25
FAQ	37	243	0	1	1	0	0	71	35
ADM	1	15	0	0	15	1	55	92	0
HOME	6007	3498	4	23	236	150	58	9421	972
NEWS	54	69	0	0	40	24	0	187	39

FIGURE 3. Objects Set: CookieId, Attribute Set: Referrer Class, Conditions Set: Access File Class

5. Test results

Starting from the PULSE database, we have performed several selections of triadic content using Toscana2TRIAS. These selections are triadic interpretations of the raw data from PULSE and have been used to test the applicability of CIRCOS to Web Usage Mining using triadic FCA. From all the performed tests, we have made a selection of those we consider representative or interesting in order to highlight navigational patterns within PULSE.

Hence, these results are intended rather as a proof of concept than a well established analysis method. In the absence of a good graphical triadic knowledge representation, a combination of several methods seems to be a reasonable approach.

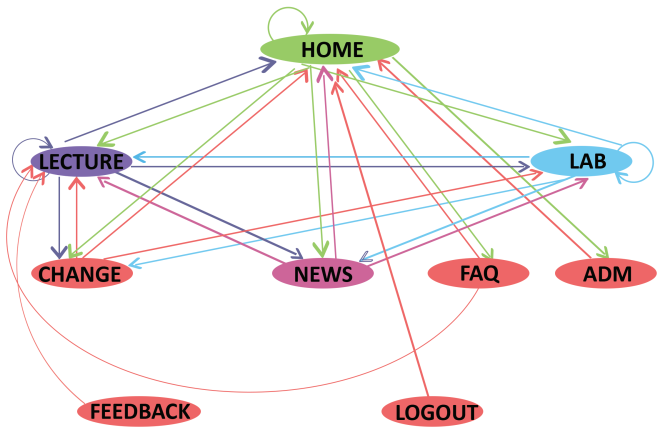


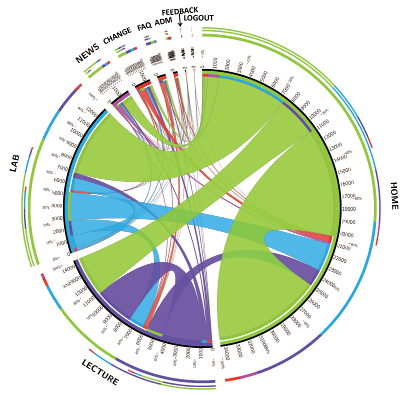
FIGURE 4. Graph visuallization of the connections between Referrer classes and Access File classes

Circular visualization of triadic content is hereby put into the discussion of the scientific community.

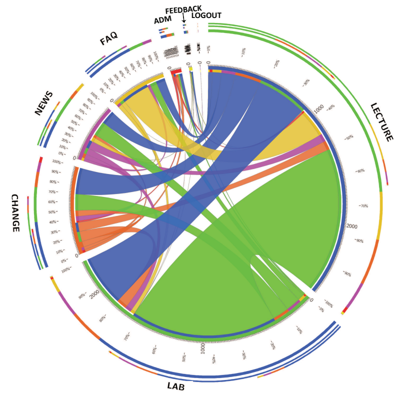
Considering the database fields of PULSE, one performed test uses Cookie ID as the object set, Referrer classes (R_class) as the attribute set and Access File classes (AF_class) as the conditions set of a tricontext. We have considered only those AF, respectively R classes which are inside PULSE, hence these two sets are the identical, but having a different semantic. The circular visualization of the obtained results, using the above mentioned algorithms is depicted in Figure 5(a).

As expected, most accesses are done from HOME. Most accesses having a webpage from the class we labeled HOME return to a webpage within the same class. This can be either the transition from login to home page as the URI does not change; the selection of the year-subject for those who have more options; or due to simply reloading the webpage. After this step, most visits go from HOME webpages to either LAB webpages, to LECTURE webpages, or to NEWS webpages. These results are reasuring as PULSE is mainly intended to serve these main purposes, i.e., to provide support for laboratory, for lectures and to post news.

The visits navigating within the LECTURE class webpages mean that they move from the lecture slides to lecture papers, explanations, and so on. These paterns of navigation inside a class is natural due to the way we constructed the classes (i.e. to contain more webpages). It can be seen within each class. In order to have a better view on what happens outside this phenomena, we eliminated them from the CIRCOS representation. We also eliminated the interactions with the HOME class as they were very well depicted in Figure 5(a). The result is depicted in Figure 5(b).



(a) Interpretation of triadic patterns arising from the triconcepts computed on Referrer classes, Access File classes and visits identified by cookieID



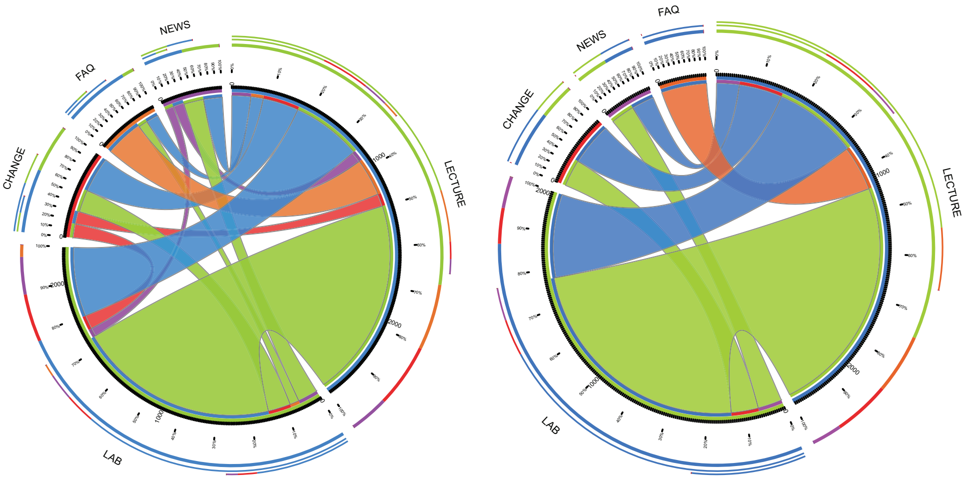
(b) Refined view from Figure 5(a) without the activity within the same class and without the HOME class

FIGURE 5

After discarding the main elements, we can deduce other patterns such as the intense activity from LAB webpages to LECTURE webpages (i.e., the biggest green ribbon). The next big trend (i.e., the biggest blue ribbon) represents the backward visits, i.e. from LECTURE webpages to LAB webpages.

A pleasant finding is to observe the use of the CHANGE facility, wich is mostly accessed from LECTURE and LAB classes. Those visits return back on red ribbons as the color coresponding to the CHANGE class is red in Figure 5(b). Another surprising finding here is that most visits from FAQ class go to LECTURE webpages, while a smaller part go to LAB webpages.

Other details are harder to be deduced from this representation. Therefore, in order to see the most important trends between the main players within PULSE we eliminated in steps, the values (i.e. number of visits) smaller than 50 and smaller than 100, because we considered they are too small to be representative, as the biggest value is larger than 1000. The results are depicted in Figure 6.



(a) Number of visits above 50

(b) Number of visits above 100

FIGURE 6. Refined view from Figure 5(b) without the classes having too few visits

The main PULSE players are the webpages from LECTURE, LAB, NEWS, CHANGE and FAQ. Figure 6(a) shows that there are more than 50 visits from LECTURE to each other class. The same is valid for the visits from LAB. For CHANGE and NEWS, the interaction is bidirectional, but only with LECTURE and LAB class, while FAQ has visits from both LAB and LECTURE, but only return on LECTURE webpages.

For the interactions with more than 100 visits the direction from the main players (i.e., LAB and LECTURE classes) is unilateral to the secondary players (i.e., CHANGE and NEWS classes).

The big number of visits from FAQ to LECTURE webpages was surprizing. After a further investigation we came to the conclusion that, the two entities were placed on the same menu on consecutive menu items. Therefore, the reasoning was that users came to FAQ webpage by mistake, trying to select the lecture menu item instead. This finding will require a design modification.

6. CONCLUSION AND FUTURE WORK

For future work, we plan to adjust certain parameters in CIRCOS configuration files. Some of the changes that can be made are: adjusting ribbon layer order, adjusting transparency of the ribbons according to the distribution of their values, hiding or removing ribbons if their values do not satisfy a certain condition, scaling cell values. Furthermore we plan to choose other

CIRCOS plot types for visualizing our data differently, such as paired-location, scattered, line, histogram, heat map, tiles, glyph.

Another issue we want to address in our future work is the large number of formal concepts extracted from data. To address this issue we plan to use an algorithm based on conditional probabilities to remove less important concepts.

Temporal data is a very important component in log access files. We plan to use Temporal Concept Analysis to study user trails and their temporal behavior in the system.

References

[1] P. Becker, J. Hereth, G. Stumme, *ToscanaJ: An Open Source Tool for Qualitative Data Analysis, Advances in Formal Concept Analysis for Knowledge Discovery in Databases.*, page 1-2. Lyon, France, (July 2002).

[2] T. Berners-Lee , W. Hall, J. Hendler, N. Shadbolt, D.J. Weitzner, *Science 313*, 769 (2006).

[3] G. Beydoun, R. Kultchitsky, G. Manasseh, *Evolving semantic web with social navigation, Expert Systems with Applications*, 32(2007), pp. 265–276.

[4] CIRCOS, a circular visualization tool, www.CIRCOS.ca

[5] A. Dieberger, *Supporting social navigation on the world wide web.*, International Journal of HumanComputer Studies, 46(6), 805825, 1997.

[6] S. Dragos, *PULSE Extended, in The Fourth International Conference on Internet and Web Applications and Services*, Venice/Mestre, Italy, May 2009, IEEE Computer Society, pp. 510–515.

[7] S. Dragos, *Why Google Analytics can not be used for educational web content*, 2011 7th International Conference on Next Generation Web Services Practices (NWeSP).

[8] S. Dragos, A.M. Beldean, *Analysing Web Usage with Force-Directed Graphs*, Studia Universitatis Babes-Bolyai Series Informatica, LVIII (2013), pp. 75–86.

[9] S. Dragos, R. Dragos, *WATEC: a Web Analytics Tool for Educational Content*, KEPT2009 Knowledge Engineering Principles and Techniques, Selected Papers, pp. 320–327.

[10] S. Dragos, C. Sacarea, *Analysing the Usage of Pulse Portal with Formal Concept Analysis*, Studia Universitatis Babes-Bolyai Series Informatica, LVII (2012), pp. 65–75.

[11] B. Ganter, R. Wille: *Formal Concept Analysis. Mathematical Foundations.* Springer, Berlin-Heidelberg-New York(1999).

[12] B. Goncalves, J. Ramasco, *Human dynamics revealed through Web analytics*, Phys. Rev. E 78, 026123 (2008).

[13] R. Jaeschke, A. Hotho, C. Schmitz, B.Ganter, G. Stumme, *TRIAS - An Algorithm for Mining Iceberg Trilattices*, Proceedings of the 6th IEEE International Conference on Data Mining ICDM 06, page 907-911. Hong Kong, IEEE Computer Society, (December 2006).

[14] The Joint Industry Committee for Web Standards (JICWEBS), *Reporting standards. website traffic.* Auditing Bureau of Circulations electronic (ABCe), Report 1, 2011.

[15] R. Kosala, H. Blockeel, *Web mining research: A survey*, ACM Sigkdd Explorations Newsletter, 2 (2000), pp. 1–15.

[16] F. Lehmann, R. Wille: *A Triadic Approach to Formal Concept Analysis*, in: Ellis, G., Levinson, R., Rich, W., Sowa, J. F. (eds.), Conceptual Structures: Applications, Implementation and Theory, vol. 954 of Lecture Notes in Artificial Intelligence, Springer Verlag, 1995.

[17] J.P. Norguet, B. Tshibasu-Kabeya, G. Bontempi, E. Zimanyi, *A Page-Classification Approach to Web Usage Semantic Analysis*, Engineering Letters, 14:1, EL 14 1 21.

[18] C. Romero, P. G. Espejo, A. Zafra, J. R. Romero, S. Ventura, *Web usage mining for predicting final marks of students that use moodle courses*, Computer Applications in Engineering Education, 21 (2013), pp. 135–146.

[19] C. Romero, S. Ventura, A. Zafra, P. D. Bra, *Applying web usage mining for personalizing hyperlinks in web-based adaptive educational systems*, Computers & Education, 53 (2009), pp. 828–840.

[20] M. Spiliopoulou, L. C. Faulstich, *Wum: a tool for web utilization analysis, in The World Wide Web and Databases*, Springer, 1999, pp. 184–203.

[21] J. Srivastava, R. Cooley, M. Deshpande, T. Pang-Ning, *Web usage mining: Discovery and applications of usage patterns from web data*, SIGKDD Explorations, 1(2), 2000.

[22] D.J. Watts, *Nature 445*, 489 (2007).

[23] B. Zhou, S. C. Hui, K. Chang, *A formal concept analysis approach for web usage mining*, in Intelligent information processing II, Springer, 2005, pp. 437–441.

[24] B. Zhou, S. C. Hui, A. C. Fong, *Web usage mining for semantic web personalization*, in Workshop on Personalization on the Semantic Web, 2005, pp. 66–72.

Babeş-Bolyai University, Faculty of Mathematics and Computer Science, 1 M. Kogălniceanu St., 400084 Cluj-Napoca, Romania

E-mail address: sanda@cs.ubbcluj.ro
E-mail address: diana.halita@ubbcluj.ro
E-mail address: csacarea@math.ubbcluj.ro
E-mail address: dianat@cs.ubbcluj.ro