Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	00000000	00000	00000000	00

Applying Triadic FCA in Studying Web Usage Behaviours

Sanda Dragoș, Diana Haliță, Christian Săcărea and Diana Troancă

Babeş-Bolyai University, Department of Computer Science, Cluj-Napoca, România sanda@cs.ubbcluj.ro, diana.halita@ubbcluj.ro, csacarea@math.ubbcluj.ro, dianat@cs.ubbcluj.ro

October 2014





イロト イポト イヨト イヨト

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

TABLE OF CONTENTS





- 3 Web Usage Mining with Triadic FCA
- 4 Test results







Why •	Prerequisites 0000000	Web Usage Mining with Triadic FCA 00000	Test results 00000000	Conclusions and Future Work

WHY

- approach: Conceptual Knowledge Processing paradigm;
- **idea**: highlight the visual part of organizing knowledge;
- **the mathematical theory**: Formal Concept Analysis (FCA);
- **purpose**: investigate web usage behavior;
- prerequisite: a previously built conceptual information system;





(コンス語) (ヨンスヨ)

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

Definition

A *formal context* is a triple (G, M, I) consisting of:

- a set of objects, G;
- a set of attributes, M;
- **a** binary relation $I \subset G \times M$, called the incidence relation.

Remark

gIm is read object g has attribute m.





・ロト ・ 雪 ト ・ 目 ト

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	00000000	00

Definition

The derivation operator in FCA *is a Galois-connection between the power set of G and the power set of M by:*

$$A' = \{ m \in M | \forall g \in A, gIm \}, A \subseteq G;$$

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





(日)

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	000000	00000	00000000	00

Definition

A triadic formal context is a quadruple (K_1, K_2, K_3, Y) , where:

- *the elements of* K_1 *are called objects;*
- the elements of K₂ are called attributes;
- *the elements of K*³ *are called conditions;*
- Y is a ternary relation, $Y \subseteq K_1 \times K_2 \times K_3$.

Remark

(g,m,b) is read object g has attribute m under condition b.





(日)

0 000€0000 000000 00	Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
	0	0000000	00000	00000000	00

Definition

A triconcept of a tricontext (K_1, K_2, K_3, Y) is a triple (A_1, A_2, A_3) with $A_i \subseteq K_i$, such that it is maximal with respect to component-wise set inclusion in satisfying $A_1 \times A_2 \times A_3 \subseteq Y$.

Remark

For a particular triconcept (A_1, A_2, A_3) :

- *A*₁ *is called the extent of the triconcept;*
- *A*₂ is called the *intent* of the triconcept;
- A_3 is called the **modus** of the triconcept.





WEB USAGE MINING

- statistics and data mining techniques are used to discover and extract useful information from web logs;
- types of sites:
 - e-commerce sites **purpose**: sell products;
 - □ e-learning sites **purpose**: offer information.
- a visit on an educational site does not apply to the heuristics used by most analytics instrument.





OUR RESEARCH

- **e-learning portal**: PULSE;
- data collection: February July 2013;
- **investigate**: patterns of web usage behavior within PULSE;
- **data fields**: the access file classes, the referrer classes, the timestamps of the system and the students login;
- ELBA & TOSCANA.





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	00000000	00000	0000000	00

ACCESS FILE CLASSES

- the request-URI represents the address of the accessed webpage along with all query information used for the actual request;
- the granularity of the accessed web pages is to fine;
- the accessed web pages have been divided into **9 classes**.



Why Pre	erequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0 00	000000	00000	0000000	00

REFERRER CLASSES

- referrer URLs represent the web page/site from which the current web page was accessed;
- the referrers which are outside of PULSE are not used in this research;



イロト イポト イヨト イヨト

WEB USAGE MINING WITH TRIADIC FCA

- Toscana2Trias allows the selection of triadic data starting from a given set of scales preprocessed with ToscanaJ;
- We used as input for the Trias algorithm:
 - Referrer class-Access File class as attribute set;
 - □ Timestamps as conditions;
 - □ Students Login as object set.





- Circos has been developed to investigate structural patterns arising in bioinformatics;
- 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.





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	00000000	00

Remark

- The XML file contains all triconcepts which can be derived from the tricontext;
- 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 × C, with numerical values.





(日)

Remark

- $\blacksquare \ (G,M,B,Y) \Leftrightarrow (G,(B,M),I), (g,(b,m)) \in I \Leftrightarrow (g,m,b) \in Y;$
- $\forall (b,m) \Rightarrow the cardinality of its extent is (b,m)';$
- C = the set of column indicators is the set obtained by projecting Y on M;
- R = the set of row indicators is the set obtaining by projecting Y on B;
- the numerical values of the table are calculated as follows: ∀(c, r) ∈ C × R, the cardinality of the extent (c,r)' is computed directly from the XML output of Trias.









Figure 3: Results for the "ar" students on the 10th week of school



Test results

Test specification

- **time**: one semester;
- **data**: group of students;

triconcepts:

- □ **objects set**: R_classes;
- attribute set: AF_classes;
- conditions: timestamps;

behavior:

- relaxed;
- intense;
- normal.

The entire set of results are posted at http:





イロト イポト イヨト イヨト

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

TEST RESULTS

the relaxed behaviour:

- □ occurs mainly during holiday;
- remark: fewer accesses and the reduced number of AF classes visited;
- □ **pattern**: HOME \Rightarrow LAB \Rightarrow LECTURE;
- □ **pattern**: HOME \Rightarrow LECTURE \Rightarrow HOME;





ヘロト 人間 とくほとくほう

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	000000	00

TEST RESULTS

the normal behaviour:

- occurs during the semester when there is no examination period or holiday;
- □ **remark**: almost all AF classes are visited;
- □ the web pages from the LAB class are more visited;





◆□▶ ◆圖▶ ◆注▶ ◆注▶

Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

TEST RESULTS

the intense behaviour:

- occurs during examination periods;
- □ **remark**: an increase number of access;
- □ the web pages from the LECTURE class are more visited;
- HOME represents a connection to the other PULSE facilities;





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

Test results - "Ar" group



Figure 4: SO: relaxed behavior

Figure 5: SO: normal behavior

Figure 6: SO: intense behavior

◆□▶ ◆圖▶ ◆注▶ ◆注▶





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	000000000	00

Test results - "ri" group



Figure 7: SO: relaxed behavior

Figure 8: SO: normal behavior

Figure 9: SO: intense behavior

◆□▶ ◆圖▶ ◆注▶ ◆注▶





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	00000000	00

Test results - "ie" group







Figure 10: WDO: relaxed behavior

Figure 11: WDO: normal behavior

Figure 12: WDO: intense behavior





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	0000000	00

Test results - "ei" group







Figure 13: WDO: relaxed behavior

Figure 14: WDO: normal behavior

Figure 15: WDO: intense behavior





CONCLUSIONS AND FUTURE WORK

- the circular visualization provides a more qualitative view on the navigational pattern, comprising more details about how and where students navigate;
- we have shown how triadic conceptual landscapes can be used for WUM and representation of user dynamics patterns;
- ADVANTAGE: the completeness of the information clustered in a concept, or determined by a derivation;
- **FURTHER RESEARCH**: apply TCA methods to describe user trails, life tracks and bundles of trails and tracks.





Why	Prerequisites	Web Usage Mining with Triadic FCA	Test results	Conclusions and Future Work
0	0000000	00000	00000000	0•

Thank you!

Q & A





ヘロア 人間ア 人間ア 人間ア