

Relevanța conținutului Web și a comportamentului utilizatorilor în analiza traficului

Diana Haliță (căs. Șotropa)

Coordonator științific:
Prof. Dr. Florian Mircea Boian

Universitatea Babeș-Bolyai, Cluj-Napoca





Direcții de cercetare

Contribuții principale

Preliminarii

Relevanța conținutului Web și a comportamentului utilizatorilor în analiza traficului

Concluzii și direcții viitoare de cercetare



Obiective

- ▶ migrare transparentă;



Obiective

- ▶ migrare transparentă;
- ▶ estimarea ratei de respingere;



Obiective

- ▶ migrare transparentă;
- ▶ estimarea ratei de respingere;
- ▶ cuantificarea nivelului de scrapiness;





Obiective

- ▶ migrare transparentă;
- ▶ estimarea ratei de respingere;
- ▶ cuantificarea nivelului de scrapiness;
- ▶ identificarea traiectoriilor în seturi de date;



Obiective

- ▶ migrare transparentă;
- ▶ estimarea ratei de respingere;
- ▶ cuantificarea nivelului de scrapiness;
- ▶ identificarea traiectoriilor în seturi de date;
- ▶ identificarea inițiatorilor de comportamente;





Obiective

- ▶ migrare transparentă;
- ▶ estimarea ratei de respingere;
- ▶ cuantificarea nivelului de scrapiness;
- ▶ identificarea traiectoriilor în seturi de date;
- ▶ identificarea inițiatorilor de comportamente;
- ▶ identificarea și formalizarea tiparelor comportamentale.



- ▶ Impactul similarității documentelor Web asupra traficului:



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;





- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;
- ▶ Tipare comportamentale în utilizarea platformelor educaționale:



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;
- ▶ Tipare comportamentale în utilizarea platformelor educaționale:
 - ▶ Web Mining;



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;
- ▶ Tipare comportamentale în utilizarea platformelor educaționale:
 - ▶ Web Mining;
 - ▶ FCA;



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;
- ▶ Tipare comportamentale în utilizarea platformelor educaționale:
 - ▶ Web Mining;
 - ▶ FCA;
 - ▶ n-FCA;



- ▶ Impactul similarității documentelor Web asupra traficului:
 - ▶ Măsuri de similaritate;
 - ▶ Web Mining;
- ▶ Tipare comportamentale în utilizarea platformelor educaționale:
 - ▶ Web Mining;
 - ▶ FCA;
 - ▶ n-FCA;
 - ▶ TCA.

Contribuții principale



- ▶ algoritmul de potrivire a perechilor de URL-uri
[Bufnea & Haliță, Studia 2013];

Contribuții principale



- ▶ algoritmul de potrivire a perechilor de URL-uri
[Bufnea & Haliță, Studia 2013];
- ▶ estimarea ratei de respingere pe baza similarității
[Haliță & Bufnea, Studia 2014];

Contribuții principale



- ▶ algoritmul de potrivire a perechilor de URL-uri
[Bufnea & Haliță, Studia 2013];
- ▶ estimarea ratei de respingere pe baza similarității
[Haliță & Bufnea, Studia 2014];
- ▶ algoritmul de cuantificare a nivelului de scrapiness;





- ▶ algoritmul de potrivire a perechilor de URL-uri
[Bufnea & Haliță, Studia 2013];
- ▶ estimarea ratei de respingere pe baza similarității
[Haliță & Bufnea, Studia 2014];
- ▶ algoritmul de cuantificare a nivelului de scrapiness;
- ▶ vizualizarea în format circular a tiparelor comportamentale folosind Circos
[Dragoș, Haliță, Săcărea & Troancă, Studia 2014]
[Dragoș, Haliță, Săcărea & Troancă, KSEM 2014]
[Dragoș, Haliță & Săcărea, SoftCOM 2015]
[Dragoș, Haliță & Săcărea, SoftCOM 2016].

Contribuții principale



- ▶ identificarea și formalizarea atractoriilor
[Dragoș, Haliță & Săcărea, KSEM 2015];



Contribuții principale



- ▶ identificarea și formalizarea atractoriilor
[Dragoș, Haliță & Săcărea, KSEM 2015];
- ▶ algoritmul de generare de tetraconcepte
[Dragoș, Haliță & Săcărea, ICCS 2016];





- ▶ identificarea și formalizarea atractoriilor
[Dragoș, Haliță & Săcărea, KSEM 2015];
- ▶ algoritmul de generare de tetraconcepte
[Dragoș, Haliță & Săcărea, ICCS 2016];
- ▶ determinarea inițiatorilor de comportamente
[Dragoș, Haliță & Troancă, AI4KM 2016];



- ▶ identificarea și formalizarea atractorilor
[Dragoș, Haliță & Săcărea, KSEM 2015];
- ▶ algoritmul de generare de tetraconcepte
[Dragoș, Haliță & Săcărea, ICCS 2016];
- ▶ determinarea inițiatorilor de comportamente
[Dragoș, Haliță & Troancă, AI4KM 2016];
- ▶ identificarea evoluției temporale a tiparelor comportamentale
[Șotropa, Studia 2016]
[Dragoș, Săcărea & Șotropa, ICFCA 2017]
[Dragoș, Șotropa & Troancă, extended paper AI4KM 2017]
[Movileanu, Săcărea & Șotropa, poster ICFCA 2017].



- ▶ analiza informațiilor (conținut, structură) prezente în Internet:



- ▶ analiza informațiilor (conținut, structură) prezente în Internet:
 - ▶ Web Content Mining;



- ▶ analiza informațiilor (conținut, structură) prezente în Internet:
 - ▶ Web Content Mining;
 - ▶ Web Structure Mining;



- ▶ analiza informațiilor (conținut, structură) prezente în Internet:
 - ▶ Web Content Mining;
 - ▶ Web Structure Mining;
 - ▶ Web Usage Mining.

Migrarea unui site Web



Necesitate

- ▶ CMS învechit;

Motivație

Provocări

Migrarea unui site Web



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

Provocări

Migrarea unui site Web



6

Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;

Provocări

Migrarea unui site Web



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;
- ▶ independența conținutului de prezentare;

Provocări

Migrarea unui site Web



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;
- ▶ independența conținutului de prezentare;
- ▶ actualizări periodice de securitate.

Provocări



Migrarea unui site Web



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;
- ▶ independența conținutului de prezentare;
- ▶ actualizări periodice de securitate.

Provocări

- ▶ expunerea conținutului site-ului Web la un URL nou;

Migrarea unui site Web



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;
- ▶ independența conținutului de prezentare;
- ▶ actualizări periodice de securitate.

Provocări

- ▶ expunerea conținutului site-ului Web la un URL nou;
- ▶ pierderea diverselor beneficii câștigate în timp;



Necesitate

- ▶ CMS învechit;
- ▶ Fructificarea noilor tehnologii.

Motivație

- ▶ întreținere facilă;
- ▶ independența conținutului de prezentare;
- ▶ actualizări periodice de securitate.

Provocări

- ▶ expunerea conținutului site-ului Web la un URL nou;
- ▶ pierderea diverselor beneficii câștigate în timp;
- ▶ creșterea numărului de vizitatori care sunt induși în eroare.

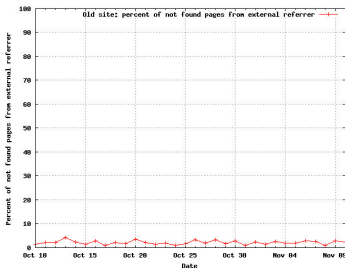
Migrarea unui site Web



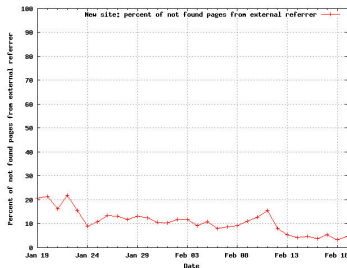
7

Soluții

- ▶ migrarea automată sau manuală;



(a) Vechiul site



(b) Noul site

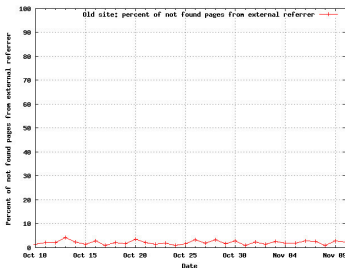
Figura: Eroare 404 generată de un referer extern

Migrarea unui site Web

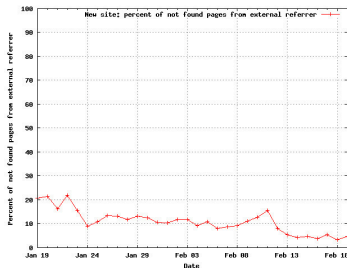


Soluții

- ▶ migrarea automată sau manuală;
- ▶ estimarea timpului necesar migrării;



(a) Vechiul site



(b) Noul site

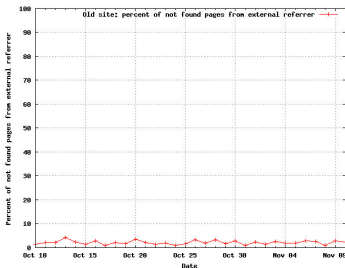
Figura: Eroare 404 generată de un referrer extern

Migrarea unui site Web

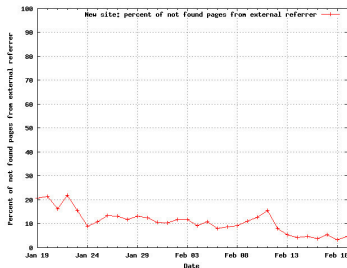


Soluții

- ▶ migrarea automată sau manuală;
- ▶ estimarea timpului necesar migrării;
- ▶ evaluarea procesului de migrare.



(a) Vechiul site



(b) Noul site

Figura: Eroare 404 generată de un referrer extern



Algoritm

- ▶ Procesare URL vechi



Algoritm

- ▶ Procesare URL vechi
- ▶ Identificare perechi de URL-uri cu conţinut similar;





Algoritm

- ▶ Procesare URL vechi
- ▶ Identificare perechi de URL-uri cu conținut similar;
- ▶ Salvare pereche + dată + similaritate;



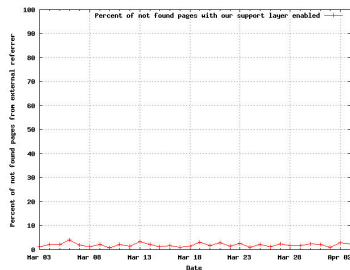
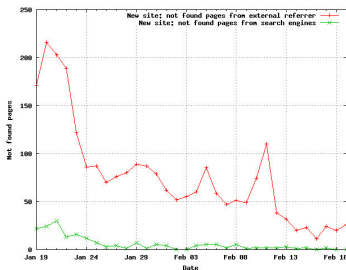
Algoritm

- ▶ Procesare URL vechi
- ▶ Identificare perechi de URL-uri cu conținut similar;
- ▶ Salvare pereche + dată + similaritate;
- ▶ Redirecționare utilizator.



Rezultate

- ▶ adaptarea motoarelor de căutare la noua structură a site-ului:



- (a) Adaptarea motoarelor de căutare vs. neadaptarea referreri-lor externi
- (b) Eroare 404 generată de referreri externi, odată cu activarea layer-ului

Figura: Noua structură a site-ului - eroarea 404 generată



Necesitate

- ▶ link-uri folosite abuziv.

Motivație

Provocări



Necesitate

- ▶ link-uri folosite abuziv.

Motivație

- ▶ creșterea page-rank-ului.

Provocări



Necesitate

- ▶ link-uri folosite abuziv.

Motivație

- ▶ creșterea page-rank-ului.

Provocări

- ▶ localizarea link-urilor abuzive.

Soluții

- ▶ analiza conținutului;

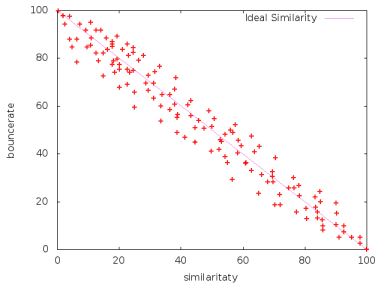


Figura: Similaritate ideală

Soluții

- ▶ analiza conținutului;
- ▶ analiza link-urilor;

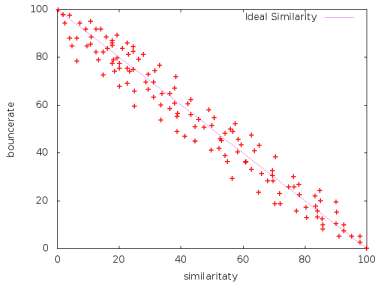


Figura: Similaritate ideală

Soluții

- ▶ analiza conținutului;
- ▶ analiza link-urilor;
- ▶ analiza comportamentului utilizatorilor.

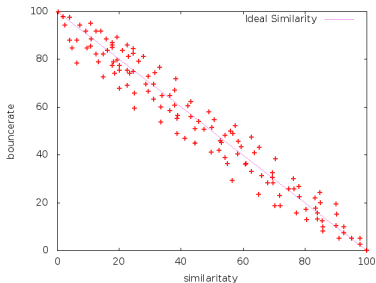


Figura: Similaritate ideală



Algoritm

- ▶ Determinarea conţinut absolut;



Algoritm

- ▶ Determinarea conținut absolut;
- ▶ Generarea triplete (pagină accesată, referrer, rată de respingere);



Algoritm

- ▶ Determinarea conținut absolut;
- ▶ Generarea triplete (pagina accesată, referrer, rată de respingere);
- ▶ Reprezentarea grafică a similarităților.

Rezultate

- ▶ cea mai bună funcție de similaritate:

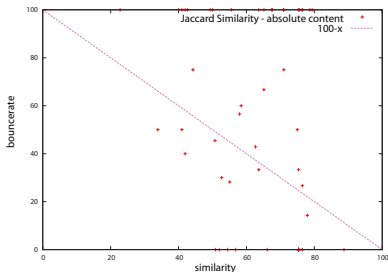


Figura: Similaritatea Jaccard



Necesitate

- ▶ direcționarea greșită a utilizatorilor (în SERP).

Motivație

Provocări



Necesitate

- ▶ direcționarea greșită a utilizatorilor (în SERP).

Motivație

- ▶ scăderea performanței motoarelor de căutare;

Provocări



Necesitate

- ▶ direcționarea greșită a utilizatorilor (în SERP).

Motivație

- ▶ scăderea performanței motoarelor de căutare;
- ▶ nemulțumirea utilizatorilor referitoare la informația găsită.

Provocări



Necesitate

- ▶ direcționarea greșită a utilizatorilor (în SERP).

Motivație

- ▶ scăderea performanței motoarelor de căutare;
- ▶ nemulțumirea utilizatorilor referitoare la informația găsită.

Provocări

- ▶ identificarea scraper-elor.



Algoritm

- ▶ Indexarea site-ului;



Algoritm

- ▶ Indexarea site-ului;
- ▶ Generarea tripletelor (link intern, link extern, similaritate);



Algoritm

- ▶ Indexarea site-ului;
- ▶ Generarea tripletelor (link intern, link extern, similaritate);
- ▶ Calcularea nivelului de scrapiness al site-ului analizat;

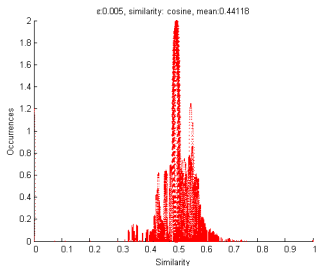


Algoritm

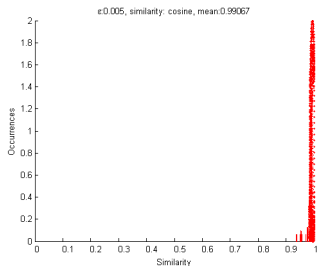
- ▶ Indexarea site-ului;
- ▶ Generarea tripletelor (link intern, link extern, similaritate);
- ▶ Calcularea nivelului de scrapiness al site-ului analizat;
- ▶ Reprezentarea grafică a tripletelor.

Rezultate

- Pentru un scraper punctele de pe grafic se regăesc în dreapta.



(a) Site C1: similaritate Cosinus



(b) Site S2: similaritate Cosinus

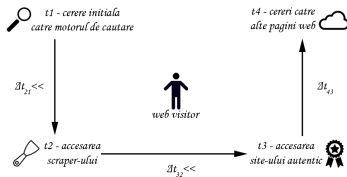
Figura: Non-Scraper vs. Scraper Site

Cuantificarea nivelului de scrapiness

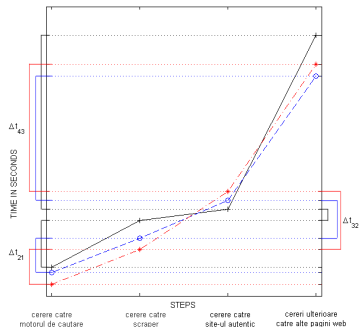


Validarea rezultatelor

- analiza log-urilor proxy;



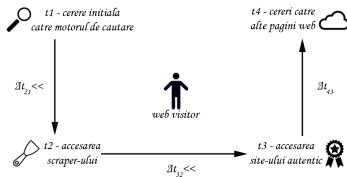
(a) Traiectorie utilizator



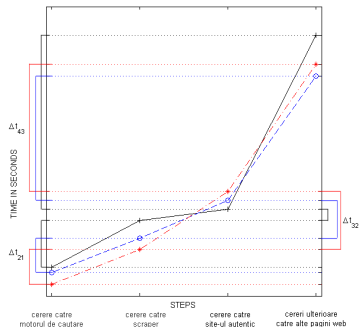
(b) Traiectorie evidențiată în logurile Squid

Validarea rezultatelor

- ▶ analiza log-urilor proxy;
- ▶ scraper-ele sunt agenți intermediari:



(a) Traiectorie utilizator



(b) Traiectorie evidențiată în logurile Squid



Motivație

- ▶ dezvoltarea metodelor de analiză de date;



Motivație

- ▶ dezvoltarea metodelor de analiză de date;
- ▶ extragerea de cunoștințe;



Motivație

- ▶ dezvoltarea metodelor de analiză de date;
- ▶ extragerea de cunoștințe;
- ▶ procesarea conceptuală a datelor;





Motivație

- ▶ dezvoltarea metodelor de analiză de date;
- ▶ extragerea de cunoștințe;
- ▶ procesarea conceptuală a datelor;
- ▶ evidențierea corelațiilor în seturile de date.



Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;

Concept Formal



Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;
- ▶ M - mulțimea atributelor;

Concept Formal





Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;
- ▶ M - mulțimea atributelor;
- ▶ $I \subseteq G \times M$ - relația de incidență.

Concept Formal



Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;
- ▶ M - mulțimea atributelor;
- ▶ $I \subseteq G \times M$ - relația de incidență.

Concept Formal

- ▶ $A \subseteq G$ - extent;



Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;
- ▶ M - mulțimea atributelor;
- ▶ $I \subseteq G \times M$ - relația de incidență.

Concept Formal

- ▶ $A \subseteq G$ - extent;
- ▶ $B \subseteq M$ - intent;



Introdusă în jurul anilor 1980 de către B. Ganter, R. Wille, P. Burmeister.

Context Formal

- ▶ G - mulțimea obiectelor;
- ▶ M - mulțimea atributelor;
- ▶ $I \subseteq G \times M$ - relația de incidență.

Concept Formal

- ▶ $A \subseteq G$ - extent;
- ▶ $B \subseteq M$ - intent;
- ▶ maximal cu proprietatea: $A \times B \subseteq I$.

Analiză Conceptuală Formală



Tabel: Context formal $\{1, 2, \dots, 10\}$ (par, impar, prim, compus, pătrat perfect)

	compus	par	impar	prim	pătrat perfect
1			x		x
2		x		x	
3			x	x	
4	x	x			x
5			x	x	
6	x	x			
7			x	x	
8	x	x			
9	x		x		x
10	x	x			

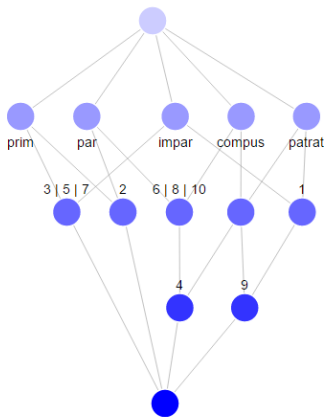


Figura: Latticea conceptelor formale



3FCA a fost introdusă de F. Lehman și R. Wille în 1995.
nFCA a fost introdusă de G. Voutsadakis în 2002.

Context Formal

- ▶ K_1, \dots, K_n - mulțimi;

Concept Formal



3FCA a fost introdusă de F. Lehman și R. Wille în 1995.
nFCA a fost introdusă de G. Voutsadakis în 2002.

Context Formal

- ▶ K_1, \dots, K_n - mulțimi;
- ▶ $R \subseteq K_1 \times \dots \times K_n$.

Concept Formal





3FCA a fost introdusă de F. Lehman și R. Wille în 1995.
nFCA a fost introdusă de G. Voutsadakis în 2002.

Context Formal

- ▶ K_1, \dots, K_n - mulțimi;
- ▶ $R \subseteq K_1 \times \dots \times K_n$.

Concept Formal

- ▶ n-uplu (A_1, \dots, A_n) ;



3FCA a fost introdusă de F. Lehman și R. Wille în 1995.
nFCA a fost introdusă de G. Voutsadakis în 2002.

Context Formal

- ▶ K_1, \dots, K_n - mulțimi;
- ▶ $R \subseteq K_1 \times \dots \times K_n$.

Concept Formal

- ▶ n-uplu (A_1, \dots, A_n) ;
- ▶ maximal cu proprietatea: $A_1 \times \dots \times A_n \subseteq R$.



TCA a fost introdusă de K. E. Wolff în 2001.

TCA

- ▶ teoria fenomenelor temporale;



TCA a fost introdusă de K. E. Wolff în 2001.

TCA

- ▶ teoria fenomenelor temporale;
- ▶ stare - concept formal al unui context formal;



TCA a fost introdusă de K. E. Wolff în 2001.

TCA

- ▶ teoria fenomenelor temporale;
- ▶ stare - concept formal al unui context formal;
- ▶ sistem - {descriere temporală, descriere a evenimentelor};



TCA a fost introdusă de K. E. Wolff în 2001.

TCA

- ▶ teoria fenomenelor temporale;
- ▶ stare - concept formal al unui context formal;
- ▶ sistem - {descriere temporală, descriere a evenimentelor};
- ▶ sistemul este într-un moment de timp în exact o stare.



Seturi de date

- ▶ platforma educațională: PULSE;

Instrumente



Seturi de date

- ▶ platforma educațională: PULSE;
- ▶ perioada de colectare a datelor: 2012-2013, 2014-2015;

Instrumente



Seturi de date

- ▶ platforma educațională: PULSE;
- ▶ perioada de colectare a datelor: 2012-2013, 2014-2015;

Instrumente

- ▶ Elba, Toscana, Toscana2TRIAS;



Seturi de date

- ▶ platforma educațională: PULSE;
- ▶ perioada de colectare a datelor: 2012-2013, 2014-2015;

Instrumente

- ▶ Elba, Toscana, Toscana2TRIAS;
- ▶ Circos;



Seturi de date

- ▶ platforma educațională: PULSE;
- ▶ perioada de colectare a datelor: 2012-2013, 2014-2015;

Instrumente

- ▶ Elba, Toscana, Toscana2TRIAS;
- ▶ Circos;
- ▶ FCA Tools Bundle;



Seturi de date

- ▶ platforma educațională: PULSE;
- ▶ perioada de colectare a datelor: 2012-2013, 2014-2015;

Instrumente

- ▶ Elba, Toscana, Toscana2TRIAS;
- ▶ Circos;
- ▶ FCA Tools Bundle;
- ▶ FACT.



Rezultate

- ▶ tipare comportamentale cantitative:



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractori;



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractori;
 - ▶ inițiatori de comportamente;



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractori;
 - ▶ inițiatori de comportamente;
 - ▶ adepți ai unor comportamente.



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractorii;
 - ▶ inițiatori de comportamente;
 - ▶ adepți ai unor comportamente.
- ▶ tipare comportamentale cantitativ-calitative:



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractori;
 - ▶ inițiatori de comportamente;
 - ▶ adepți ai unor comportamente.
- ▶ tipare comportamentale cantitativ-calitative:
 - ▶ comportament timpuriu;



Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractorii;
 - ▶ inițiatori de comportamente;
 - ▶ adepți ai unor comportamente.
- ▶ tipare comportamentale cantitativ-calitative:
 - ▶ comportament timpuriu;
 - ▶ comportament uzual;

Rezultate

- ▶ tipare comportamentale cantitative:
 - ▶ relaxat;
 - ▶ normal;
 - ▶ intens.
- ▶ tipare comportamentale calitative:
 - ▶ atractori;
 - ▶ inițiatori de comportamente;
 - ▶ adepți ai unor comportamente.
- ▶ tipare comportamentale cantitativ-calitative:
 - ▶ comportament timpuriu;
 - ▶ comportament uzual;
 - ▶ comportament întârziat.



Atratori

- ▶ tipare comportamentale la care utilizatorii aderă în timp ce utilizează platforma educațională;

Clasificare



Atratori

- ▶ tipare comportamentale la care utilizatorii aderă în timp ce utilizează platforma educațională;

Clasificare

- ▶ structurați: atractor comportamental, atractor amplu, atractor de navigare, atractor critic



Atratori

- ▶ tipare comportamentale la care utilizatorii aderă în timp ce utilizează platforma educațională;

Clasificare

- ▶ structurați: atractor comportamental, atractor amplu, atractor de navigare, atractor critic
- ▶ nestructurați: atractor comportamental, atractor popular, atractor critic



Atractor comportamental

- ▶ scală conceptuală pe o granulă de timp;

Atractor amplu



Atractor comportamental

- ▶ scală conceptuală pe o granulă de timp;
- ▶ reflectă obiceiurile de navigare ale utilizatorilor;

Atractor amplu



Atractor comportamental

- ▶ scală conceptuală pe o granulă de timp;
- ▶ reflectă obiceiurile de navigare ale utilizatorilor;

Atractor amplu

- ▶ scală conceptuală pe o granulă de timp



Atractor comportamental

- ▶ scală conceptuală pe o granulă de timp;
- ▶ reflectă obiceiurile de navigare ale utilizatorilor;

Atractor amplu

- ▶ scală conceptuală pe o granulă de timp
- ▶ reflectă comportamentele ample ale utilizatorilor;

Atrator de navigare

- ▶ structură conceptuală intenționată;

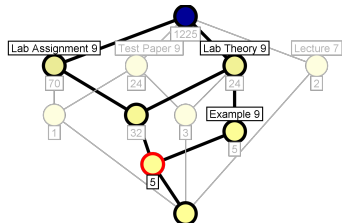
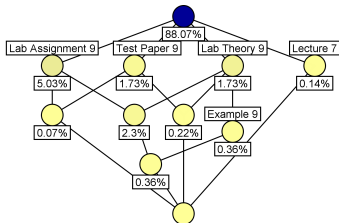
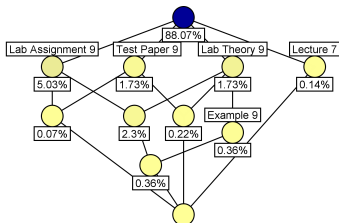


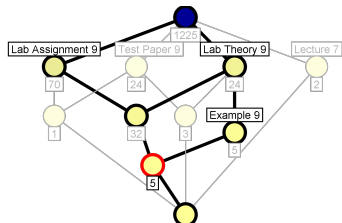
Figura: Atratorul educațional pentru laboratorul 9

Atractor de navigare

- ▶ structură conceptuală intenționată;
- ▶ reflectă scopurile educaționale ale instructorului.



(a) Materiale de învățare

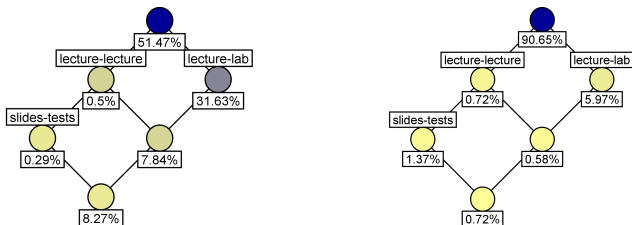


(b) Materiale de laborator (teorie, exemple, teme atribuite)

Figura: Atractorul educațional pentru laboratorul 9

Atrator popular

- ▶ scală conceptuală pe granule de timp;

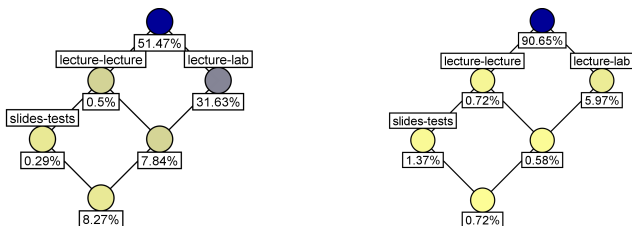


(a) ramificații conținând una din paginile etichetate în noduri
(b) ramificații consecutive care încep și se termină cu una din paginile etichetate

Figura: Lanțuri de pagini care conțin în ramificații pagini referitoare la materialele de curs

Atractor popular

- ▶ scală conceptuală pe granule de timp;
- ▶ reflectă tipare de navigare neprevăzute;

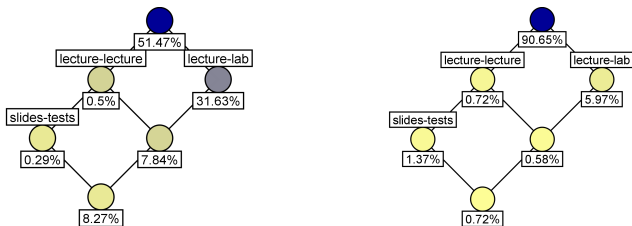


(a) ramificații conținând una din paginile etichetate în noduri
(b) ramificații consecutive care încep și se termină cu una din paginile etichetate

Figura: Lanțuri de pagini care conțin în ramificații pagini referitoare la materialele de curs

Atractor popular

- ▶ scală conceptuală pe granule de timp;
- ▶ reflectă tipare de navigare neprevăzute;
- ▶ comportamentul de ramificare a lanțului de pagini vizitate.

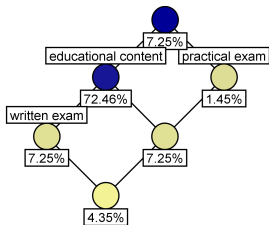


(a) ramificații conținând una din paginile etichetate în noduri
(b) ramificații consecutive care încep și se termină cu una din paginile etichetate

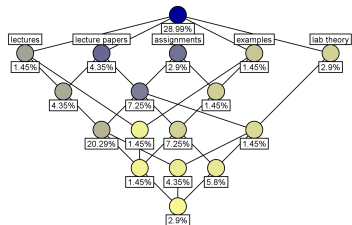
Figura: Lanțuri de pagini care conțin în ramificații pagini referitoare la materialele de curs

Atractor critic

- ▶ atractor popular împreună cu o granulă de timp din *CP*;



(a) Ce pagini sunt vizitate?

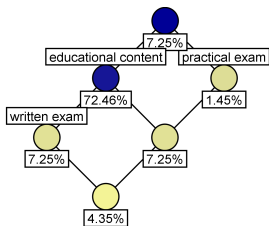


(b) Ce conținut educațional este vizitat?

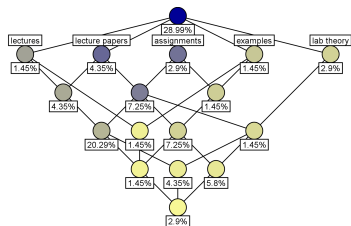
Figura: Comportament calitativ: ce pagini vizitează studenții în timpul pregătirii pentru examen?

Atractor critic

- ▶ atractor popular împreună cu o granulă de timp din CP ;
- ▶ scală conceptuală care reflectă comportamentul utilizatorului determinat de evenimentele critice.



(a) Ce pagini sunt vizitate?



(b) Ce conținut educațional este vizitat?

Figura: Comportament calitativ: ce pagini vizitează studenții în timpul pregătirii pentru examen?



Algoritm

- ▶ generalizarea algoritmului Trias;



Algoritm

- ▶ generalizarea algoritmului Trias;
- ▶ cuboidele se completează cu a 4-a dimensiune ținând cont de condiția de maximalitate.



Utilizator: K , clase: LA-I

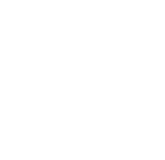
- ▶ $([H, K, O, T], [2], [2], [LA-I])$
- ▶ $([E], [1, 2, 3], [6], [LA-I])$
- ▶ $([K], [2], [2], [LA-I, WE-LA-I])$
- ▶ $([B, H, T], [3], [2], [LT-LA-I])$
- ▶ $([H], [3], [2], [LT-LA-I, LT-LA])$
- ▶ $([K], [3], [3], [LT-LA-I, L-LA-I, LT-LA, L-LA, L-WE-LA])$
- ▶ $([O], [3], [2], [L-LA-I])$
- ▶ $([B], [9], [7], [L-PE-LA-I])$
- ▶ $([K], [16], [16], [LP-LPs-PE-LA-I, LP-LPs, L-LPs, L-WE, Ls-L-LP])$



Inițiatori de comportamente

- ▶ utilizatori care inițiază anumite comportamente;

Adepti ai unor comportamente





Inițiatori de comportamente

- ▶ utilizatori care inițiază anumite comportamente;

Adepti ai unor comportamente

- ▶ utilizatori care preiau comportamente



Obiecte	Atribute	Conditii	Stări 1	Stări 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

Tabel: Eșantion de concepte pentadice grupate după comportament, inițiator și săptămâna în care a apărut comportamentul



Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;



Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;



Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;
- ▶ mulțimea obiectelor actuale: $\Pi \subseteq P \times G$;



Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;
- ▶ mulțimea obiectelor actuale: $\Pi \subseteq P \times G$;
- ▶ mulțimea granulelor temporale în care utilizatorii au avut activitate:



Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;
- ▶ mulțimea obiectelor actuale: $\Pi \subseteq P \times G$;
- ▶ mulțimea granulelor temporale în care utilizatorii au avut activitate:
 - ▶ $Early_p = \{g \in G | (p, g) \in \Pi, g < n\}$;



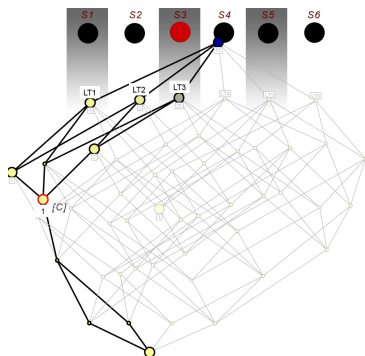
Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;
- ▶ mulțimea obiectelor actuale: $\Pi \subseteq P \times G$;
- ▶ mulțimea granulelor temporale în care utilizatorii au avut activitate:
 - ▶ $Early_p = \{g \in G \mid (p, g) \in \Pi, g < n\}$;
 - ▶ $Common_p = \{g \in G \mid (p, g) \in \Pi, n \leq g \leq n + 2\}$;

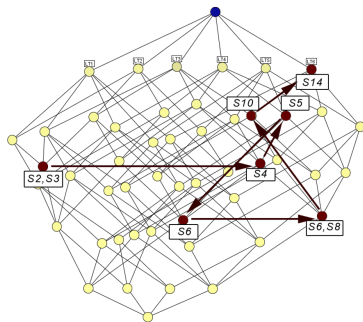


Comportamente cantitativ-calitative: timpuriu, uzual și întârziat

- ▶ mulțimea utilizatorilor: $P = \{st_1, st_2, \dots, st_x\}$;
- ▶ mulțimea granulelor temporale: $G = \{w_1, w_2, \dots, w_y\}$;
- ▶ mulțimea obiectelor actuale: $\Pi \subseteq P \times G$;
- ▶ mulțimea granulelor temporale în care utilizatorii au avut activitate:
 - ▶ $Early_p = \{g \in G | (p, g) \in \Pi, g < n\}$;
 - ▶ $Common_p = \{g \in G | (p, g) \in \Pi, n \leq g \leq n + 2\}$;
 - ▶ $Late_p = \{g \in G | (p, g) \in \Pi, g > n + 2\}$.

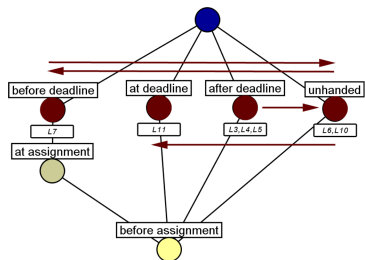


(a) Comportament timpuriu

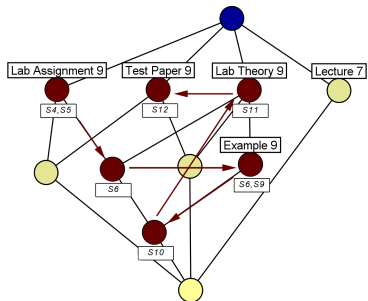


(b) Traietorii

Figura: Focus pe Lab Theory, primele 6 săptămâni



(a) Predarea laboratoarelor



(b) Gradul de aderare la atractorul educațional 9

Figura: Traectoria utilizatorului C



- ▶ definirea unor metode cu posibile aplicații în clasificarea și ierarhizarea site-urilor Web;



- ▶ definirea unor metode cu posibile aplicații în clasificarea și ierarhizarea site-urilor Web;
- ▶ identificarea metodelor de navigare și explorare a seturilor de date n-adice;



- ▶ definirea unor metode cu posibile aplicații în clasificarea și ierarhizarea site-urilor Web;
- ▶ identificarea metodelor de navigare și explorare a seturilor de date n-adice;
- ▶ identificarea tiparelor comportamentale prin metode cantitative și calitative;



- ▶ definirea unor metode cu posibile aplicații în clasificarea și ierarhizarea site-urilor Web;
- ▶ identificarea metodelor de navigare și explorare a seturilor de date n-adice;
- ▶ identificarea tiparelor comportamentale prin metode cantitative și calitative;
- ▶ utilizarea metodelor TCA împreună cu tehnica construirii scalelor.



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;
- ▶ detectarea tiparelor comportamentale utilizând FCA Relațional și Tipare Structurate;



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;
- ▶ detectarea tiparelor comportamentale utilizând FCA Relațional și Tipare Structurate;
- ▶ aplicarea metodelor identificate pe alte seturi de date:



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;
- ▶ detectarea tiparelor comportamentale utilizând FCA Relațional și Tipare Structurate;
- ▶ aplicarea metodelor identificate pe alte seturi de date:
 - ▶ seturi de date medicale;



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;
- ▶ detectarea tiparelor comportamentale utilizând FCA Relațional și Tipare Structurate;
- ▶ aplicarea metodelor identificate pe alte seturi de date:
 - ▶ seturi de date medicale;
 - ▶ seturi de date colectate de pe site-uri comerciale;



- ▶ detectarea comportamentelor similare utilizând măsuri de similaritate FCA;
- ▶ detectarea tiparelor comportamentale utilizând FCA Relațional și Tipare Structurate;
- ▶ aplicarea metodelor identificate pe alte seturi de date:
 - ▶ seturi de date medicale;
 - ▶ seturi de date colectate de pe site-uri comerciale;
- ▶ determinarea dinamicii utilizatorilor în site-uri de tip spam.

Mulțumesc pentru atenție!



Lista de publicații I

- ▶ D. Bufnea, **D. Haliță**, *A server-side support layer for transparent web content migration*, Proceedings of the International Conference on Knowledge Engineering, Principles and Techniques, Studia Universitatis Babeș-Bolyai Informatica, Volume LVIII, no. 3, 2013, pp. 78-89.
- ▶ **D. Haliță**, D. Bufnea, *A study regarding inter domain linked documents similarity and their consequent bounce rate*, Studia Universitatis Babeș-Bolyai Informatica, Volume LIX, no. 1, 2014, pp. 83-91.
- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, D. Troancă, *An FCA grounded study of user dynamics through log exploration*, Studia Universitatis Babeș-Bolyai Informatica, Volume LIX, No. 2, 2014, pp. 82-97.

Lista de publicații II

- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, D. Troancă, *Applying Triadic FCA in Studying Web Usage Behaviors*, Proceedings of the 7th International Conference on Knowledge Science, Engineering and Management, Volume 8793, 2014, pp. 73-80.
- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, *Attractors in Web Based Educational Systems. A Conceptual Knowledge Processing Grounded Approach*, Proceedings of the 8th International Conference on Knowledge Science, Engineering and Management, Volume 9403, 2015, pp. 190-195.
- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, *Behavioral pattern mining in web based educational system*, Proceedings of the 23rd International Conference Software, Telecommunications and Computer Networks, IEEE Catalog Number CFP1587A-ART, 2015, pp. 215-219.

Lista de publicații III

- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, *Analysing the Effect of Changing the Educational Methods by Using FCA*, Proceedings of the 24th International Conference Software, Telecommunications and Computer Networks, IEEE Catalog Number CFP1687A-ART, 2016, pp. 243-247.
- ▶ S. Dragoș, **D. Haliță**, C. Săcărea, *Distilling Conceptual Structures from Weblog Using Polyadic FCA*, Proceedings of the 22nd International Conference on Conceptual Structures, Volume LNCS 9717, 2016, pp. 151-159.
- ▶ S. Dragoș, **D. Haliță**, D. Troancă, *Investigating Trend-setters in E-learning Systems using Polyadic Formal Concept Analysis and Answer Set Programming* Proceedings of the 4th International Workshop on Artificial Intelligence for Knowledge Management, 2016, pp. 42-48.

Lista de publicații IV

- ▶ **D. Haliță**, C. Săcărea, *Is FCA suitable to improve Electronic Health Record Systems?*, Proceedings of the 24th International Conference Software, Telecommunications and Computer Networks, IEEE Catalog Number CFP1687A-ART, 2016, pp. 248-252.
- ▶ **D. Șotropa**, *A FCA Strategy for Improving Web-Based Learning Sites* Studia Universitatis Babeș-Bolyai Informatica, Volume LXI, No. 2, 2016, pp. 110-125.
- ▶ S. Dragoș, C. Săcărea, **D. Șotropa**, *An investigation of user behavior in educational platforms using Temporal Concept Analysis*, Proceedings of the 14th International Conference on Formal Concept Analysis, 2017, to be published.

Lista de publicații V

- ▶ S. Dragoș, **D. Șotropa**, D. Troancă, *Highlighting Trend-setters in Educational Platforms by Means of Formal Concept Analysis and Answer Set Programming*, under review for the 4th AI4KM Extended Revised Papers.
- ▶ B. Movileanu, C. Săcărea, **D. Șotropa**, *FACT - A tool for Temporal Formal Concept Analysis*, under review for the ICFCA 2017 poster session.

Bibliografie I

- ▶ L. Becchetti, C. Castillo, D. Donato, S. Leonardi, *Web Spam Detection: link-based and content-based techniques*, Proceedings of the final workshop of the European Integrated Project Dynamically Evolving, Large Scale Information Systems, Volume 222, 2008, pp. 99-113.
- ▶ 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*, Proceedings of the workshop FCAKDD of the 15th European Conference on Artificial Intelligence Advances in Formal Concept Analysis for Knowledge Discovery in Databases, Volume 426, 2002, pp. 1-2.
- ▶ R. Cooley, B. Mobasher, J. Srivastava, *Web mining: Information and pattern discovery on the world wide web*, Proceedings of the 9th IEEE International Conference on Tools with Artificial Intelligence, 1997, pp. 558-567.

Bibliografie II

- ▶ S. Dragoş, *PULSE Extended, Proceedings of the Fourth International Conference on Internet and Web Applications and Services*, 2009, pp. 510-515.
- ▶ G. Gan, M. Chaoqun, W. Jianhong, *Data Clustering: Theory, Algorithms, and Applications*, ASA-SIAM Series on Statistics and Applied Probability, 2007.
- ▶ B. Ganter, R. Wille, *Formal Concept Analysis. Mathematical Foundations*, 1999.

Bibliografie III

- ▶ R. Jaeschke, A. Hotho, C. Schmitz, B. Ganter, G. Stumme, *Trias - An Algorithm for Mining Iceberg Trilattices*, Proceedings of the IEEE International Conference on Data Mining, 2006, pp. 907-911.
- ▶ L. L. Kis, C. Săcărea, D. Troancă, *FCA Tools Bundle - a Tool that Enables Dyadic and Triadic Conceptual Navigation*, Proceedings of the 4th International Workshop FCA4AI "What can FCA do for Artificial Intelligence?", 2016, pp. 43-45.
- ▶ M. Krzywinski, J. Schein, I. Birol, J. Connors, R. Gascoyne, D. Horsman, S.J. Jones, M. A. Marra, *Circos: an information aesthetic for comparative genomics*, Genome research, Volume 19, no. 9, 2009, pp. 1639-1645.

Bibliografie IV

- ▶ F. Lehmann, R. Wille, *A Triadic Approach to Formal Concept Analysis*, Conceptual Structures: Applications, Implementation and Theory, Volume LNAI 954, 1995.
- ▶ M. Najork, *Detecting Spam web Pages through Content Analysis*, International World Wide web Conference Comittee, 2006, pp. 83-92.
- ▶ G. Voutsadakis, *Polyadic concept analysis*, Order, Volume 19, no. 3, 2002, pp. 295-304.
- ▶ K. E. Wolff, *Temporal concept analysis*, International Workshop on Concept Lattices-Based Theory, Methods and Tools for Knowledge Discovery in Databases, 2001, pp. 91-107.