

UTILIZAREA WEB MINING-ULUI ÎN APLICAȚII E-COMMERCE

Claudia Elena Dinucă
Facultatea de Economie și
Administrarea Afacerilor,
Universitatea din Craiova
clauely4u@yahoo.com

Rezumat :

În zilele noastre, Web-ul reprezintă o parte importantă din viața umană. Web-ul este considerat cel mai bun spațiu pentru a dezvolta afaceri. În ziua de azi, companiile mari își regândesc afacerea folosind Internetul pentru a-și îmbunătăți afacerea. Afacerea realizată pe Web oferă oportunități pentru potențialii clienți sau parteneri unde se găsesc produsele solicitate și companii specifice. Prezența unui site al unei companii oferă câteva avantaje în comparație cu existența unei afaceri într-un spațiu fizic deoarece elimină barierele de timp și spațiu. Pentru a se diferenția în cadrul economiei pe Internet, întreprinderile învingătoare au realizat că e-commerce înseamnă mai mult decât simple tranzacții de cumpărare/vânzare, strategiile corespunzătoare fiind cheia succesului pentru a îmbunătăți puterea de competiție. Una dintre tehnicile eficiente folosite în acest scop este data mining. Data mining este procesul extragerii cunoștințelor interesante din date. Web mining reprezintă folosirea tehnicilor data mining pentru extragerea informațiilor din datele de pe web. Acest articol prezintă cele trei componente ale web mining: web usage mining, web structure mining și web content mining.

CUVINTE CHEIE: e-commerce, web mining, web content mining, web structure mining, web usage mining.

1. Web mining

Web-ul a devenit foarte popular în ultimul deceniu, aducând astfel o puternică platformă de diseminare a informațiilor și extragere a cunoștințelor, precum și analiza informațiilor, în zilele noastre Web-ul fiind cunoscut drept un depozit mare de date conținând o mare varietate de tipuri de date, precum și cunoștințe de baza, în care sunt ascunse informații despre Web [1].

USING WEB MINING IN E-COMMERCE APPLICATIONS

Claudia Elena Dinucă
Faculty of Economics and Business
Administration,
University of Craiova
clauely4u@yahoo.com

Abstract :

Nowadays, the web is an important part of our daily life. The web is now the best medium of doing business. Large companies rethink their business strategy using the web to improve business. Business carried on the Web offers the opportunity to potential customers or partners where their products and specific business can be found. Business presence through a company web site has several advantages as it breaks the barrier of time and space compared with the existence of a physical office. To differentiate through the Internet economy, winning companies have realized that e-commerce transactions is more than just buying / selling, appropriate strategies are key to improve competitive power. One effective technique used for this purpose is data mining. Data mining is the process of extracting interesting knowledge from data. Web mining is the use of data mining techniques to extract information from web data. This article presents the three components of web mining: web usage mining, web structure mining and web content mining.

KEYWORDS: e-commerce, web mining, web content mining, web structure mining, web usage mining.

1. Web mining

The Web has become very popular over the last decade, bringing a strong platform for information dissemination and retrieval and analysis of information, today the Web being known as a large data warehouse containing a wide variety of data and knowledge base, in which information about Web knowledge are hidden [1].

Utilizatorii întâmpină probleme datorită volumului mare de informații care este într-o continuă creștere. În mod special, utilizatorii Web-ului întâmpină probleme în găsirea informațiilor corecte datorită preciziei scăzute și rechemării scăzute a paginilor. De exemplu, dacă un utilizator dorește să extragă anumite informații folosind motorul de căutare Google, acesta va oferi nu numai conținuturi Web în legătură cu acest subiect, ci și o serie de informații irelevante, așa numitele pagini zgomot, determinând astfel dificultăți utilizatorilor în obținerea informațiilor necesare [1]. Toate acestea determină o provocare cercetătorilor din domeniul Web-ului pentru descoperirea metodelor de management și extragere eficientă a informațiilor de pe Web.

Web-ul reprezintă un canal critic de promovare și comunicare a imaginii unei companii. Cercetarea pe Web are drept scop descoperirea informațiilor din surse masive de date folosind tehnici data mining. Cercetarea pe Web se ocupă cu reprezentări de date semi-structurate și heterogene, precum date textuale, structuri hyperlink și informații de utilizare pentru a îmbunătăți calitatea serviciilor oferite de diferite aplicații Web. Astel de aplicații acoperă o largă categorie de subiecte, precum obținerea de conținuturi Web dorite, extragerea și analizarea comunităților Web, profilul utilizatorilor, customizarea prezentărilor Web în concordanță cu preferințele utilizatorilor [1]. Astfel, site-urile de comerț electronic reprezintă importante canale de vânzări. Este foarte important să se folosească metode data mining pentru a analiza datele din cadrul activităților executate de vizitatori pe aceste site-uri web.

În general, e-commerce și e-business (în continuare denumit în continuare e-commerce) au permis tranzacții on-line. De asemenea, generarea de baze de date de mari dimensiuni în timp real de date nu a fost niciodată mai ușoară.

Data mining (DM) are ca scop dominant, generarea de informații non-evidente încă utile pentru factorii de decizie de la baze de date foarte mari. Aceste forme, la rândul lor, sunt rezultatul punerii în aplicare a tehnicilor de modelare sofisticate din diverse domenii de statistici, inteligența artificială, de gestionare a bazelor de date

Cercetarea Web mining-ului se împarte în trei categorii: mining-ul conținutului Web (Web content mining), mining-ul structurii Web (Web structure

Users experience problems due to the large volume of information that is constantly growing. In particular, Web users have problems finding the correct information due to low precision and low recall page. For example, if a user wants to retrieve some information using Google search engine, it will provide not only Web contents dealing with this topic, but a series of irrelevant information, so-called noise pages, resulting in difficulties for users in obtaining necessary information [1]. All this leads to a challenge of researchers to discover the web management methods and effective extraction of information from the web.

The Web is a critical channel of communication and promoting a company image. The research aims to discover Web information from massive sources of data using data mining techniques. The research on the Web deals with representations of semi-structured and heterogeneous data, such as textual information, hyperlink structure and receive information of use to improve the quality of services provided by different Web applications. These kind of applications cover a wide category of subjects, such as obtaining the desired Web content, retrieval and analysis of Web communities, user profiles, customized Web presentations in accordance with user preferences [1]. E-commerce sites are important sales channels. E-commerce has changed the face of most business functions in competitive enterprises. It is very important to use data mining methods to analyze data from the activities carried out by visitors to these websites.

In general, e-commerce and e-business (henceforth referred to as e-commerce) have enabled on-line transactions and generating large-scale real-time data has never been easier.

Data mining (DM) has as its dominant goal, the generation of non-obvious yet useful information for decision makers from very large databases. These forms in turn are the result of the implementation of sophisticated modeling techniques in various fields of statistics, artificial intelligence, database management.

Web mining research is divided into three categories: mining the Web content (Web content mining), mining the web structure (web structure

mining) și mining-ul utilizării Web (Web usage mining).

Web content mining este procesul de extragere a cunoștințelor din conținutul documentelor și din descrierea acestora. Web structure mining este procesul obținerii de cunoștințe din modul de organizare a web-ului și din legăturile între paginile web. În cazul web usage mining sunt analizate informațiile despre vizitele paginilor web salvate în fișierele de log ale serverelor de internet în vederea descoperirii de modele interesante necunoscute anterior și posibil utile. Web usage mining este descris ca aplicarea de tehnici data mining pe logurile de acces web pentru optimizarea site-ului web în beneficiul utilizatorilor.

2. Web Content Mining

În ultimii ani, creșterea WWW a depășit toate așteptările. Astăzi sunt câteva miliarde de documente HTML, imagini și alte fișiere multimedia disponibile pe Internet și numărul lor este într-o continuă creștere. Extragerea informațiilor interesante a devenit o sarcină foarte dificilă luând în considerare varietatea impresionantă de pe Web.

Web content mining este procesul de extragere a informațiilor folositoare din conținutul documentelor Web. Conținutul Web-ului constă în câteva tipuri de date precum date text, imagini, audio și video, înregistrări structurate precum liste sau tabele și hyperlink-uri.

Cercetarea a fost făcută în scopul acoperirii diferitelor părți ale web content mining, îmbogățirea modului în care site-urile web sunt prezentate utilizatorului, îmbunătățirea calității rezultatelor căutării și extragerea conținuturilor interesante ale paginilor web.

Tehnici de regăsire a informațiilor sunt luate în calcul când avem de a face cu date nestructurate și semi-structurate din diverse domenii ale statisticii, inteligenței artificiale, gestionarea bazelor de date și grafică pe calculator

Vederea bazei de date a datelor Web este utilizată pentru a avea o mai bună gestionare a informațiilor și interogărilor pe Web, procesul de extragere întotdeauna încearcă să deducă structura site-ului pentru a transforma un site într-o bază de date.

3. Web Structure Mining (Web Linkage Mining)

Web Linkage transmite informațiile despre modul în care paginile diferite sunt asociate împreună și formează imensul Web [1].

mining) and mining the Web use (Web usage mining).

Web content mining is the process of extracting knowledge from documents and content description. Web structure mining is the process of obtaining knowledge from the organization of the Web and the links between Web pages.

Web usage mining analyzes informations about visited Web page that saved in log files of Internet servers in order to discover interesting patterns previously unknown and potentially useful. Web usage is described as mining applying data mining techniques on Web access logs to optimize web site for users.

2. Web Content Mining

In recent years, the growth of the World Wide Web exceeded all expectations. Today there are several billions of HTML documents, pictures and other multimedia files available via the internet and the number is still rising. Retrieving the interesting contents has become a very difficult task taking into consideration the impressive variety of the Web.

Web content mining is the process of extracting useful information from the content of Web documents. Web content consists of several types of data such as text data, images, audio or video, structured records such as lists or tables and hyperlinks.

The research was done to cover different parts of web content mining, enrichment the way how websites are presented to the user, improving the quality of search results and extracting interesting web pages contents.

Information retrieval techniques are taken to deal with the unstructured and semi-structured data, the diverse fields of statistics, artificial intelligence, database management and computer graphics.

Database view of Web data is used to have the better information management and querying on the Web, the mining always tries to infer the structure of the Website to transform a Website to become a database.

3. Web Structure Mining(Web Linkage Mining)

Web Linkage conveys the information how various pages are associated together and formed the huge Web [1].

Explorarea legăturilor web relevă structura de bază ascunse în hyperlink-uri și le folosește pentru mai multe aplicații web, cum ar fi cele de căutare. Web structure mining încearcă să descopere legătura structurii de hyperlink-uri la nivelul inter-documentului pentru a genera un rezumat structural despre site-ul și pagina web.

Deci, în web structure mining avem următoarele direcții:

- bazate pe hyperlinkurile, clasificarea paginilor Web și generarea informațiilor.
- descoperirea structurii documentului Web în sine.
- descoperirea naturii ierarhiei sau rețeaua de hyperlink-uri în site-ul unui anumit domeniu.

Web structure mining descrie organizarea conținutului web în cazul în care structura este definită prin hyperlink-uri între pagini și comenzile HTML de formatare într-o pagină. Înțelegerea relației dintre conținutul și structura site-ului este utilă pentru a menține o imagine de ansamblu despre site-uri web. În acest fel, poate fi indicat dacă o pagină se potrivește în conținutul structurii legăturilor sale și se identifică teme care se extind peste mai multe pagini conectate. Link-urile către un document indică popularitatea documentului, în timp ce legăturile ce ies dintr-un document indică bogăția sau varietatea de subiecte cuprinse în document.

În 1998, doi algoritmi importanți au fost proiectați pe bază de hyperlink-uri: PageRank și HITS. Ambele PageRank [2] și HITS [3], au la origine analiza rețelelor de socializare. Ei exploatează structura hyperlinkurilor paginilor web pentru a acorda ranguri paginilor în funcție de gradul lor de prestigiu sau autoritate.

Algoritmul Page Rank a fost creat în perioada 1997-1998. Pe baza acestui algoritm funcționează cel mai de succes motor de căutare pe internet, Google. Page Rank își are originea în analiza rețelelor sociale și are ca principiu acordarea unui rang fiecărei pagini web în funcție de câte legături dinspre alte pagini web conduc la pagina respectivă.

Se presupune că avem n pagini web, fiecare pagină i conținând un număr O_i de linkuri spre alte pagini web. Se notează cu A matricea de adiacență asociată web-ului privit ca un graf orientat $G=(V, E)$ în care paginile sunt vârfuri iar linkurile dintre pagini sunt arcele grafului.

Matricea de adiacență asociată grafului va avea

$$\text{elementele : } A_{ij} = \begin{cases} \frac{1}{O_i}, & \text{dacă } (i, j) \in E \\ 0, & \text{dacă } (i, j) \notin E \end{cases} \quad (1)$$

Web linkage mining reveals the underlying structure hidden in the hyperlinks and utilizes them for more web applications such as search. Web structure mining tries to discover the link structure of the hyperlinks at the inter-document level to generate structural summary about the Website and Web page.

In the web structure mining we have the following directions :

- based on the hyperlinks, categorizing the Web pages and generated the information.
- discovering the structure of Web document itself.
- discovering the nature of the hierarchy or network of hyperlinks in the Website of a particular domain.

Web structure mining describes the organization of the content of the web where structure is defined by hyperlinks between pages and HTML formatting commands within a page. Understanding the relationship between contents and the structure of the website is useful to keep an overview about websites. In this way, it can be indicated whether a page fits in the content of its link structure and identify topics which span over several connected pages. Links pointing to a document indicates the popularity of the document, whereas link coming out of a document indicate the richness or the variety of topics covered in the document.

In 1998, two most important hyperlink based algorithms were designed: PageRank and HITS. Both PageRank[2] and HITS[3] were originated from the social network analysis. They exploit the hyperlink structure of the Web to rank pages according to their degree of prestige or authority.

Page Rank algorithm was created in 1997-1998. The most successful Internet search engine, Google works based on this algorithm. Page Rank is rooted in social network analysis and is basically providing a ranking of each web page depending on how many links from other Web pages lead to that page.

Assume that we have n Web pages, each page i containing a number of O_i links to other websites. Note A the adjacency matrix associated with the website seen as a directed graph $G = (V, E)$ where pages are vertices and links between pages are arcs of the graph. Associated graph adjacency matrix will have elements

$$A_{ij} = \begin{cases} \frac{1}{O_i}, & \text{if } (i, j) \in E \\ 0, & \text{if } (i, j) \notin E \end{cases} \quad (1)$$

Plecând cu un vector inițial de probabilități, utilizând o matrice stocastică ireductibilă și aperiodică, conform Teoremei Ergodice a lanțurilor Markov se obține un șir de vectori de probabilități convergent către o stare unică de echilibru

$$P_1 = A^T P_0, \quad (2)$$

$$P_k = A^T P_{k-1}, \quad (3)$$

$$\lim_{k \rightarrow \infty} P_k = P. \quad (4)$$

Vectorul de probabilități obținut ne va da rangul paginilor web.

Pentru a putea aplica Teorema Ergodică, matricea de adiacență este transformată pentru a îndeplini condițiile de ireductibilitate și aperiodicitate.

Formula

$$P(i) = (1-d) + d \sum_{(j,i) \in E} \frac{P(j)}{O_j}, \quad (5)$$

ne dă rangul paginii i, unde P(i) este rangul paginii i iar d este un factor de atenuare ce ia valori între 0 și 1.

Algoritmul pseudocod pentru calculul rangurilor paginilor web este:

PageRank

$$P_0 \leftarrow \frac{e}{n}$$

$$k \leftarrow 1$$

repetă

$$P_k \leftarrow (1-d)e + dA^T P_{k-1};$$

$$k \leftarrow k + 1;$$

$$\text{până când } \|P_k - P_{k-1}\|_1 < \varepsilon$$

afișează P_k

Unde e este vectorul cu toate elementele 1, ε este pragul de acuratețe solicitat iar 1-norma unui vector este dată de suma elementelor componente.

4. Web Usage Mining

Web usage mining este partea cea mai relevantă din punct de vedere al marketing-ului, deoarece explorează căile de navigare și comportamente din timpul vizitei în cadrul unui site web al unei companii. Metodele de extragere a regulilor de asociere sunt folosite pentru obținerea corelațiilor dintre diferitele pagini vizitate în timpul unei sesiuni. Regulile de asociere secvențiale sau modele de serii de timp pot fi folosite pentru a analiza datele utilizate dintr-un site web luând în

Starting with an initial vector of probabilities, using a stochastic irreducible and aperiodic matrix, according to Ergodic Theorem of Markov chains, we obtain a range of vectors of probabilities converges to a unique state of equilibrium

$$P_1 = A^T P_0, \quad (2)$$

$$P_k = A^T P_{k-1}, \quad (3)$$

$$\lim_{k \rightarrow \infty} P_k = P. \quad (4)$$

Vector of probabilities obtained will give us web pages rank.

To apply Ergodic Theorem adjacency matrix is transformed to meet conditions of irreducibility and aperiodicity.

The next formula

$$P(i) = (1-d) + d \sum_{(j,i) \in E} \frac{P(j)}{O_j}, \quad (5)$$

gives the rank of page i, P(i), and d is a mitigating factor that takes values between 0 and 1.

The pseudocode algorithm for calculating the ranking of web pages is:

PageRank

$$P_0 \leftarrow \frac{e}{n}$$

$$k \leftarrow 1$$

repeat

$$P_k \leftarrow (1-d)e + dA^T P_{k-1};$$

$$k \leftarrow k + 1;$$

$$\text{until } \|P_k - P_{k-1}\|_1 < \varepsilon$$

display P_k

Where e is a vector with all elements 1, ε is called the threshold of accuracy and 1-norm of a vector is given by the sum of constituents.

4. Web Usage Mining

Web usage mining is the most relevant part in terms of marketing because it explores ways to navigate and conduct during a visit to the website of a company. Methods for extracting association rules are useful for obtaining correlations between the various pages visited during a browsing session. Sequential association rules and time series models can be used to analyze used data from a Web site taking in account a temporal dynamics using the

calcul o dinamică temporală a utilizării site-ului. Web usage mining se bazează în principal pe analiza secvenței de pagini vizitate în cadrul unei sesiuni date, numită analiza clicurilor web.

Informații despre comportamentul de cumpărare al vizitatorilor se pot obține în cadrul site-urilor web de comerț electronic analizând clicurile web.

În cazul web usage mining sunt analizate informațiile despre vizitele paginilor web salvate în fișierele de log ale serverelor de internet în vederea descoperii de modele interesante necunoscute anterior și posibil utile. Web usage mining este descris ca aplicarea de tehnici de data mining pe logurile de acces web pentru optimizarea sitului web în beneficiul utilizatorilor.

Prin click-stream se înțelege o secvență de pagini web vizualizate de un utilizator, paginile fiind afișate câte una pe rând la un moment dat.

În momentul în care un vizitator accesează un site web, serverul reține toate acțiunile făcute de vizitator într-un fișier jurnal(log file). O sesiune a utilizatorului descrie succesiunea paginilor vizualizate de utilizator în momentul unei perioade de logare pe web, pagini din cadrul mai multor site-uri.

O sesiune(sesiune a server-ului) reprezintă totalitatea paginilor unui site web din cadrul unei sesiuni de utilizator, cunoscută și sub denumirea de vizita.

Fiecare click de mouse corespunde unei cereri de pagina web, succesiunea click-urilor corespunde astfel succesiunii link-urilor.

Analiza fluxului click-urilor succesive poate fi folosită pentru a înțelege cea mai posibilă cale de navigare într-un site web cu scopul previziunii online a paginilor pe care un vizitator le va accesa dată fiind succesiunea link-urilor(calea) pe care a urmat-o până atunci. Paginile vor fi create dinamic, ideea vizualizării unei pagini particulare putând fi astfel extinsă la ideea vizualizării unei clase particulare de pagini sau tipuri de pagini, clasa fiind definită prin meta informațiile continute în header.

Clickstream sau clickpaths reprezintă drumul click-urilor de mouse pe care fiecare vizitator îl face în timp ce navighează într-un site. Inițial, analiza clickstream determină de unde a venit vizitatorul site-ului, calea urmată în cadrul paginilor web pentru care se face analiza, timpul petrecut pe fiecare pagina, precum și când și unde se deplasează vizitatorul după ce părăsește sesiunea

site. Web usage mining is mainly based on sequence analysis of pages visited during a given session, analyzing web clicks.

Information about the purchasing behavior of visitors can be taken from the e-commerce sites by analyzing web clicks.

In web usage mining, there are analyzed information on web pages visits that are saved in log files of Internet servers in order to discover the previously unknown and potentially interesting useful patterns. Web usage mining is described as applying data mining techniques on Web access logs to optimize web site for users interest.

The click-stream means a sequence of Web pages viewed by a user, pages are displayed one by one at a time.

When a visitor accesses a website, the server retains all the actions taken by visitors in a log file. A user session describes the sequence of pages viewed by the user during a period of logging on the web pages from several sites.

A session (server session) represents all visited pages from a website within a user session, known as the visit.

Each click of the mouse corresponds to a web page request, the sequence of clicks corresponds to such sequence links.

Successive clicks flow analysis can be used to understand the possible way to navigate a web site for online prediction of the pages a visitor will access knowing the sequence of links (paths) that has followed one before. Pages are dynamically created, the idea of viewing a particular page can be extended to the idea of viewing a particular class of pages or page types, as defined by the meta class information contained in the header.

Clickstream or path is clickpaths mouse clicks that each visitor makes while browsing a site. Initially, sequence analysis of Clickstream determine from where it came the site visitors, the path followed in the web pages for which the analysis is done, time spent on each page, and when and where the visitor travels after leaving the

din cadrul site-ului. Statistica de ansamblu folosind informații cumulate despre clickstream ajută la generarea de rapoarte web analitice cum sunt frecvența vizitelor și durata pe care persoanele o petrec pe un site particular și activitatea lor, precum frecvența vizitelor și transformarea acestor vizite în activități comerciale.

Analiza click-urilor web este metoda de adunare și analizare a datelor despre modul în care vizitatorii site-ului îl traversează. Clickstream reprezintă ordinea în care o persoană vizitează paginile.

În principal, sunt două tipuri de analiza a click-urilor și anume:

- analiza traficului
- analiza e-commerce

Analiza traficului implică analiza drumului pe care utilizatorul îl urmează prin site bazându-se pe datele adunate la nivel de server folosind date din analiza clicurilor. Analiza traficului, de asemenea, înregistrează numărul de câte ori s-au încărcat paginile în timpul navigării.

Analiza e-commerce folosește date din clickstream pentru a determina eficacitatea de marketing a site-ului prin cuantificarea comportamentului utilizatorului în timp ce vizitează site-ul practic înregistrând traducerea navigării vizitatorului pe site în tranzacții de vânzare. Analiza e-commerce este indicator al gradului de comoditate al utilizatorului în folosirea interfeței formelor, cosului de cumpărături, plății etc.

Analiza clickstream a devenit un aspect foarte important al afacerilor online și a publicității. Deoarece înregistrează modul în care utilizatorii se mută de la o pagină la alta, se poate observa cum potențialii clienți au ajuns la aceste pagini și dacă ei staționează pe aceste pagini ale afacerii electronice sau merg mai departe. Analiza tipului de utilizator pentru un anumit site web permite realizarea unui alt design al site-ului pentru a îmbunătăți gradul de satisfacție al clientului. Analiza click-urilor este importantă atât pentru afacerile electronice cât și pentru determinarea locurilor ideale pentru publicitate. După cum este specificat în articolul *Clickstream Analysis: Both a Business and an Aid in Advertising*, Stephanie Mattingly[16] analiza click-urilor reprezintă un ajutor semnificativ în publicitate. Această analiză poate determina paginile sursă de unde s-a ajuns la paginile site-ului curent, precum și faptul dacă campaniile de publicitate pe care le rulează sunt de succes sau nu.

session of the site. Overall statistics using aggregate information about Clickstream helps us generate analytical reports such as frequency of visits and time that people spend on a particular site and their activity, and frequency of visits and turning those visits into commercial activities.

Analysis of web clicks is the method of gathering and analyzing data about how it crosses the site visitors. Therefore, Clickstream is the order in which the people visit web pages.

Basically, there are two types of analysis of clicks, namely:

- traffic analysis (traffic analysis)
- e-commerce (e-commerce analysis)

Traffic analysis involves analyzing road traffic that the user follows the site which is based on data collected at the server level analysis using data from clicks. Traffic analysis also records the number of times users loaded pages while browsing.

Analysis of e-commerce use Clickstream data to determine the marketing effectiveness of the site by quantifying user behavior while actually visiting the site visitor browsing the site recording the translation in sales transactions. Analysis of e-commerce is an indicator of the degree of user convenience in using the interface forms, shopping cart, payment etc.

Clickstream analysis has become a very important aspect of online business and advertising. Because they record how users move from one page to another, it can be seen how potential customers have come to these pages and if it stays on the pages of this electronic business or go on. By making the user profile for a specific site, clickstream analysis has become a type of analysis used for a given Web site which allows another site design to improve customer satisfaction. Analysis of clicks is important for electronic business and to determine the ideal places for advertising. As specified in the article *Clickstream Analysis: Both a Business and an Aid in Advertising*, Stephanie Mattingly[16] clicks analysis is a significant help in advertising. This analysis can determine the source pages from where it was reached the current web site, and whether the advertising campaigns we run are successful or not.

Se poate determina dacă există alte site-uri cu legături către paginile curente, care sunt aceste site-uri, precum și numărul de vizitatori care au fost direcționați către paginile curente datorită acestor legături. Analiza clicurilor poate arăta dacă anunțurile de pe un site conduc vizitatorii afară din acest site, astfel în loc să câștige clienți, acestea direcționează utilizatorii către alte locuri pentru a cumpăra.

Scopul principal în ajutorul publicității este de a arăta o conexiune între anunțuri care sunt plasate pe internet și cumpăraturile făcute de pe site-ul web. Analiza clicurilor ajută și la îmbunătățirea site-ului web prin personalizarea acestuia.

Analiza coșului de cumpărături presupune determinarea tuturor produselor care sunt cumpărate împreună de către cumpărători. Analiza coșului de cumpărături este una dintre aplicațiile intuitive ale regulilor de asociere. De exemplu, se poate descoperi că clienții cumpără lapte și pâine împreună și se pot determina chiar tipurile specifice de lapte care sunt cumpărate împreună cu anumite tipuri de pâine, de exemplu pâine integrală cu lapte de soia. Acestea precum și alte multe reguli încă nedescoperite pot fi folosite pentru a maximiza profitul prin utilizarea lor la realizarea unor campanii de marketing de succes și prin modul de realizare a design-ului magazinului cel real cât și cel virtual. În cazul exemplului cu laptele și pâinea, poate că vânzătorul nu oferă reducere la ambele produse, ci doar pentru unul, dar celalalt produs trebuie așezat strategic cât mai departe de produsul anterior pentru a îmbunătăți traficul în cadrul magazinului real sau virtual, creând astfel posibilitatea ca acest cumpărător să mai dorească să cumpere și alte produse văzute între timp.

5. Concluzii

E-commerce a schimbat fața majorității funcțiilor ale afacerii în întreprinderi competitive. E-commerce face clienții disponibili în aproape toate industriile. Clienții pot cumpăra non-stop și companiile pot oferi aplicații self-service și de distribuție de produse și servicii către clienți. E-commerce oferă noi forme de condițiile de piață care schimbă radical jocul. Prin afaceri electronice este redefinit fiecare proces al afacerii tradiționale: de la dezvoltarea de noi produse și managementul relației cu clienții până la achiziționarea resurselor umane, materiale și componente de achiziții publice. Scopul aplicației e-commerce este de a furniza informațiile necesare pentru utilizatorul final, și, de asemenea, atrage clienții pentru a reveni mai târziu. Serviciile oferite utilizatorilor cu

It can be determined if there are other sites with links to current pages, which are the sites and the number of visitors which were directed to the current due to these links. Clickstream analysis can show whether clicks ads on a site run out of this site so instead of leads customers, they direct users to other places to buy.

The main purpose in advertising help is to show a connection between the ads that are placed on the Internet and purchases made on website. The analysis also helps to improve click website by customizing it.

Shopping basket analysis involves determining all products that are purchased together by customers. Shopping basket analysis is one of the applications of association rules intuition. For example, you may discover that customers buy milk and bread together and can even cause specific types of milk are bought with certain types of bread, such as bread with soy milk. These and many other undiscovered rules can be used to maximize profits by helping to achieve successful marketing campaigns and by the manner of the design store, the real and the virtual one. In the milk and bread example, the seller cannot offer discount on both products, but only one, but other items must be placed in such a way as far as previous product to improve traffic in real or virtual store, creating the possibility that this buyer was also willing to buy other products seen in the meantime.

5. Conclusions

E-commerce has changed the face of most business functions in competitive enterprises. E-commerce solutions make customers available in almost all industries. Customers can buy non-stop and companies can offer self-service applications and distribution of products and services to customer. E-commerce offers new forms of market conditions which radically change the game. Through electronic business is redefined every process in the traditional business: from the development of new products and customer relationship management to purchasing human resources and materials and components procurement. The goal of e-commerce application is to provide necessary information to the end user, and also attract clients to come later. The services

privire la funcționalitatea, ușurința de utilizare (de navigare și organizare mai ușor de înțeles) și puterea de răspuns al cererii, integritatea informațiilor este esențială pentru percepția utilizatorului. În cazul în care utilizatorul poate găsi aplicația utilă atunci este o șansă pentru a reveni pentru a face afaceri noi.

Internetul și e-commerce este o cheie importantă pentru succesul viitor al oricărei organizații, care oferă oportunități uriașe și piață de desfacere în întreaga lume. Proiectele e-commerce sunt sortite eșecului din cauza neînțelegerii noilor norme din mediu e-economiei, prin urmare, faptul că nu țin pas competiției.

Pentru a se diferenția în economia pe Internet, companiile trebuie să înțeleagă că companiile câștigătoare înseamnă mai mult decât simple tranzacții de cumpărare / vânzare, strategii adecvate fiind cheia pentru îmbunătățirea puterii competitive. Acest lucru poate fi realizat prin utilizarea tehnicilor de web mining pe date istorice de la activităților de comerț electronic. Web mining poate fi împărțit după cum a fost precizat mai sus în trei categorii: Web Content Mining, Web Structure Mining și Web Usage Mining.

Data mining, ca aplicație în e-commerce este o tehnologie revoluționară care poate aduna informații într-o manieră automată și poate construi modele folosite pentru a prezice deciziile de cumpărare a clienților cu o acuratețe remarcabilă.

În încheiere, unele direcții de investigat pot fi considerate în funcție de conținuturile de e-commerce care au fost discutate aici, urmând efectuarea de cercetări suplimentare cu privire la aceste domenii specifice.

Referințe :

- [1] Guandong Xu, Zanchun Yhang, Lin Li, Web Mining and Social Networking Techniques and Applications, USA:Springer, 2011
- [2] J. Palau, M. Montaner, B. Lopez and J.L. de la Rosa. Collaboration analysis in the recommender system using social networks. In CIA, pages 137-151,2004
- [3] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. In Proc. Of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '98), pages 668-677, 1998
- [4] Zaiane O., Han J.: WebML: Querying the World Wide Web for resources and knowledge. In: Workshop on Web Information and Data Management WIDM98, Bethesda, 1998, 9-12.

provided to users on the functionality, ease of use (navigation and organization easier to understand) and response strength of the application, integrity of information is critical to user perception. If the user finds application useful then it is a chance to return to make new business.

The Internet and e-commerce is an important key to future success of any organization, offering huge opportunities and market outlets worldwide. E-commerce projects are doomed to failure because of misunderstanding the new rules of e-economy environment, thus failing to step into the competition.

To differentiate into the Internet economy, companies must realize that winning e-commerce means more than simple transactions of purchase/sale, the appropriate strategies being the key to improving competitive power. This can be done by using web mining techniques on historical data from e-commerce activities. Web mining can be divided as was stated above in three categories: Web content mining, Web structure mining and Web usage mining. Data mining as applied to e-commerce is a breakthrough technology that can gather information in an automated fashion and build models used to predict customer purchasing decisions with remarkable accuracy.

To conclude, some directions may be taken depending on the e-commerce contents that were discussed here conducting further research on those particular areas.

References:

- [1] Guandong Xu, Zanchun Yhang, Lin Li, Web Mining and Social Networking Techniques and Applications, USA:Springer, 2011
- [2] J. Palau, M. Montaner, B. Lopez and J.L. de la Rosa. Collaboration analysis in the recommender system using social networks. In CIA, pages 137-151,2004
- [3] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. In Proc. Of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA '98), pages 668-677, 1998
- [4] Zaiane O., Han J.: WebML: Querying the World Wide Web for resources and knowledge. In: Workshop on Web Information and Data Management WIDM98, Bethesda, 1998, 9-12.

[5] Yang, Q., Zhang, H.H. and Li, I.T, 2001. Mining Web logs for prediction models in www caching and prefetching, Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 26-29, pp. 473-478.

[6] Yan, T.W., Jacobsen, M. Garcia-Molina, H. and Dayal, U., 1996, Knowledge Discovery from users web page navigation, Seventh International Workshop on Research Issues in Data Engineering, Birmingham, England, aprilie 7-8, pp 20-29.

[7] Berry, M., Linoff, G.: Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley and Sons, Chichester (1997)

[8] Dunham, M.H.: Data Mining : Introductory and Advanced Topics. Prentice Hall, Pearson Education Inc. (2003)

[9] Prinzie, A. , Van den Poel, D.: Investigating Purchasing Patterns for Financial Services using Markov, MTD and MTDg Models. In: Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 03/213 (2003)

[10] Agrawal, R., Srikant, R.: Mining sequential patterns, *International Conference on Data Engineering(ICDE'95)*, Taipei, Taiwan, pp. 3-14, martie 1995.

[11] Schechter, S. Krishnan, M. Smith, M.D. :*Using path profiles to predict http request*, *Seventh International World Wide Web Conference*, Brisbane, Australia, pp. 457-467, Aprilie, 1998.

[12] Nong, Y.: The handbook of Data Mining, Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey, 2003.

[13] Vercellis, C. : *Business Intelligence: Data Mining and Optimization for Decision Making*,UK: John Wiley & Sons, 2009

[14] Jiawei Han, Micheline Kamber: *Data Mining Concepts and Techniques* Second Edition, USA: Elsevier, 2006

[15] Guandong Xu, Zanchun Yhang, Lin Li, Web Mining and Social Networking Techniques and Applications, USA:Springer, 2011

[16] Stephanie Mattingly ,*Clickstream Analysis: Both a Business and an Aid in Advertising*, Disponibil online:
<http://cseweb.ucsd.edu/~paturi/cse91/Presents/smattingly.pdf>

[5] Yang, Q., Zhang, H.H. and Li, I.T, 2001. Mining Web logs for prediction models in www caching and prefetching, Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, San Francisco, CA, USA, August 26-29, pp. 473-478.

[6] Yan, T.W., Jacobsen, M. Garcia-Molina, H. and Dayal, U., 1996, Knowledge Discovery from users web page navigation, Seventh International Workshop on Research Issues in Data Engineering, Birmingham, England, aprilie 7-8, pp 20-29.

[7] Berry, M., Linoff, G.: Data Mining Techniques for Marketing, Sales and Customer Support, John Wiley and Sons, Chichester (1997)

[8] Dunham, M.H.: Data Mining : Introductory and Advanced Topics. Prentice Hall, Pearson Education Inc. (2003)

[9] Prinzie, A. , Van den Poel, D.: Investigating Purchasing Patterns for Financial Services using Markov, MTD and MTDg Models. In: Working Papers of Faculty of Economics and Business Administration, Ghent University, Belgium 03/213 (2003)

[10] Agrawal, R., Srikant, R.:Mining sequential patterns, *International Conference on Data Engineering(ICDE'95)*, Taipei, Taiwan, pp. 3-14, martie 1995.

[11] Schechter, S. Krishnan, M. Smith, M.D. :*Using path profiles to predict http request*, *Seventh International World Wide Web Conference*, Brisbane, Australia, pp. 457-467, Aprilie, 1998.

[12] Nong, Y.: The handbook of Data Mining, Lawrence Erlbaum Associates, Publishers Mahwah, New Jersey, 2003.

[13] Vercellis, C. : *Business Intelligence: Data Mining and Optimization for Decision Making*,UK: John Wiley & Sons, 2009

[14] Jiawei Han, Micheline Kamber: *Data Mining Concepts and Techniques* Second Edition, USA: Elsevier, 2006

[15] Guandong Xu, Zanchun Yhang, Lin Li, Web Mining and Social Networking Techniques and Applications, USA:Springer, 2011

[16] Stephanie Mattingly ,*Clickstream Analysis: Both a Business and an Aid in Advertising*, Available online:
<http://cseweb.ucsd.edu/~paturi/cse91/Presents/smattingly.pdf>