

**Приложение на извличането на знания от Интернет източници при управлението на връзките с клиентите**

Доц. д-р Снежана Динева Сълова

**Application of Web Mining in Customer Relationship Management**

Assoc. prof. Snezhana Dineva Sulova, PhD

Abstract

*The accumulation of more and more information on the Internet makes the global network a suitable source of new knowledge. Thus Web Mining technology and its application in customer relationship management are studied.*

*Keywords: web mining, customer relationship management (CRM), internet knowledge*

За проследяване на цялата история на развитието на взаимоотношенията с клиентите, координиране на многостранните връзки с тях и осъществяване на ефективен клиентски ориентиран маркетинг се използват системите за управление на връзките с клиентите (CRM). Те подпомагат фирмите при осъществяване на обслужването и както и при цялостната работа с клиентите. От съществено значение за работата на CRM системите е информационната база и натрупаните в нея данни. За поддръжка на ефективен маркетинг, продажби и обслужване на клиентите обаче, считаме, че може да се използва и наличната информация в Интернет. Събраните данни за клиентите при регистрация или чрез формуляри и анкети не винаги са достоверни и пълни, тъй като много от клиентите от липса на достатъчно мотивация или от съображения за сигурност избират да не попълват предоставяните им въпросници.

Анализите, базирани на данни от Интернет макар и да не са изключително прецизни, в много случаи са носители на полезна информация за потребителските посещения, за интересите на потребителите, за тяхното поведение. Те имат голямо значение за персонализирането на услугите и за насочването на потребителите към информация, която би била максимално полезна за тях. *В тази връзка цел на настоящата статия е да се покаже приложението на технологиите за извличането на знания от Интернет източници за събиране на допълнителни данни за клиентите и използването им за подобряване на дейностите по управление на взаимоотношенията с тях.*

Световната мрежа Интернет, с бурните си темпове на развитие обхваща все повече потребители. Според последни данни броят им е над 3,2 млрд. души<sup>1</sup>. Това води до непрекъснатото увеличаване на обема на информационни ресурси в мрежата. Доказано е, че глобалната мрежа удвоява обема си на всеки 5,32 години<sup>2</sup>. Публикациите в мрежата са динамични, обновяват се непрекъснато, появяват се актуални гледни точки за проблемите и това ги прави ценен източник на нови знания. Същевременно, обаче, необходимо е да се отбележи, че Интернет ресурсите са твърде различни по своя характер. Основните източници на информация са уеб сайтовете, които включват в своето съдържание информация под формата на текст, таблици, изображения, мултимедия (аудио, видео). Подари този хетерогенен характер на информацията извличането на знания от тях е сложен и динамичен процес.

Извличането на знания от Интернет ресурси се нарича Web Mining (WM). Един от първите изследователи на този процес е Ециони, който през 1996 г. го определя като използване на DM техниките за изследване и извличане на информация от Интернет документи и услуги<sup>3</sup>. По-късно това определение се допълва от като към него анализаторите вече

---

1 Internet World Stats, <http://www.internetworldstats.com/stats.htm> (30.09.2015)

2Кръстева, Н. Законът на Мур е в сила и за Интернет ресурсите, [http://cio.bg/4380\\_zakonat\\_na\\_mur\\_e\\_v\\_sila\\_i\\_za\\_internet\\_resursite](http://cio.bg/4380_zakonat_na_mur_e_v_sila_i_za_internet_resursite), (30.09.2015)

включват и техниките за изследване и анализ на данните за използваемостта на уеб ресурсите и дефинират WM като “разкриване на закономерности, зависимости, шаблони в съдържанието, структурата и начина на използването на уеб документите, които показват как да се приложи извличането на знания от данни (Data Mining – DM) върху уеб базирани данни”<sup>4</sup>.

Въпреки, че както се вижда от определенията извличането на данни от Интернет източници води началото си от DM и използва същите методи, необходимо е да се отбележи, че двата процеса не са идентични. Основната причина за това е, че традиционно извличането на знания работи на базата на структурирани данни, съхранявани в складове от данни, реляционни бази от данни, електронни таблици или текстови данни в таблична форма, а информацията при Интернет източниците по своето същество има полу или не структуриран вид и се нуждае от предварителна обработка за да се сведе до вид подходящ за анализ.

В зависимост от основните типове Интернет ресурси, които се използват в процеса на WM, в литературата<sup>5</sup> са дефинирани три основни вида извличане на знания (вж. фиг. 1):



Фиг. 1 Видове WM

- извличане на знания от съдържанието на Интернет източниците (Web Content Mining - WCM);
- извличане на знания от структурата на Интернет източниците (Web Structured Mining - WSM);
- извличане на знания от информацията за използваемостта на Интернет източниците (Web Usage Mining - WUM).

Процесът на извличане на знания от Интернет ресурси се осъществява в следните основни етапи (вж. фиг. 2):

- търсене на подходящи Интернет ресурси;
- предварителна обработка на ресурсите с цел структуриране на данните;
- откриване на модели в данните;
- анализ;
- получаване на нови знания.

3 Etzioni, O. The World Wide Web: quagmire or gold mine?, Communications of the ACM, no. 11/1996, p. 65-68.

4 Markov, Z. and D. Larosed. Data Mining the Web Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons, New Jersey, 2007.

5 Cooley, R. Mobasher, B. and Srivastave, J., Web Mining: Information and Pattern Discovery on the World Wide Web, <http://maya.cs.depaul.edu/classes/ect584/papers/cms-tai.pdf>, (9.12.2011).

Markov, Z. and D. Larosed. Data Mining the Web Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons, New Jersey, 2007.

Kosala, R. and H. Blockeel. Web Mining Research: A Survey. <http://facweb.cs.depaul.edu/mobasher/classes/ect584/papers/kosala.pdf>, (1.10.2015).

Cronin, B. Annual review of information science and technology, Vol. 38, Information today Inc., 2004.



Фиг. 2 Етапи на процеса на извличане на знания от Интернет ресурси

При търсене на необходимата информация в мрежата се използват търсещите машини, които обхождат милиони документи и индексират думите в уеб източниците, запазвайки си индексите за последващо използване в други търсения.

Следващият етап от процеса на WM, който е и от първостепенно значение е предварителната обработка на информацията, събрана от Интернет източниците. За осъществяването му няма ясно дефинирани правила, защото начинът на обработка на Интернет източниците зависи до голяма степен от техния вид. Ако се анализира текст от публикуван в Интернет документ или от уеб страница се използват алгоритмите за извличане на съдържание от неструктурирани и полуструктурирани документи. Известно е, че в текстовите документи има голямо количество думи и не е целесъобразно всички те да се подлагат на анализ, тъй като не всички са носители на полезна информация, затова от текстовете първо се премахват излишните думи, извличат се фрази и словосъчетания, съставят се релации. Ако Интернет ресурсът е от друг вид, например лог файл от Интернет сървър се прилагат специални техники за извличане на данните от него и преобразуването им в таблична форма.

Следва прилагане на методите за извличане на знания, като класификация, регресия, клъстеризация, търсене на асоциативни правила и последователности с цел да се открият модели в данните и на базата на тях да се извърши анализ и тълкуване на резултатите. Извличането на знания помага да се разбере защо са се случили съответните събития, да се открият тенденции, да се направят асоциация с подобни събития.

Представеният процес на извличане на знания от Интернет ресурси може да се приложи за набиране на данни и извличане на нови знания за всички бизнес дейности, които се осъществяват посредством използването на приложения работещи в Интернет среда. Анализите на информация от Интернет източници намират приложение в области, където на са достатъчни само статистическите и аналитичните методи и изгражданите чрез тях модели. Една от тези области е управлението на връзките с клиентите.

В динамичната пазарна среда, в която работят фирмите управлението на връзките с клиентите се определя като интегрирана стратегия по продажбите, маркетинга и услугите. Осъществяването ѝ позволява по-добре да се интегрира клиентът в организацията. Фирмите се стремят да получават максимално възможната информация за своите клиенти и техните

потребности и изхождайки от нея си съставят ясна фирмена стратегия за спечелване на нови клиенти и за задържане на съществуващите.

Емпирично е установено, че в Интернет бизнеса е много по-трудно да се задържи един клиент, като с изключение на най-известните и сериозни онлайн магазини (напр. Амазон), процентът на връщащите се клиенти е нисък. От друга страна, най-висока възвръщаемост има именно от постоянните клиенти. Ето защо основна част от фазите и задачите на CRM приложенията са именно насочени към подпомагане на дейностите по запазване на съществуващите клиенти и извличане на максимална полза от тях.

В тази връзка за извършване на по-задълбочени анализи, които да са в основата на една ориентирана към клиента стратегия считаме, че е целесъобразно да се обработва и използва и наличната информация в Интернет. Събраните данни в базата от данни за клиентите не са достатъчни за проучване и предвиждане на нуждите на сегашните и на потенциалните клиенти. В много от случаите анализите, базирани на извлечени от Интернет данни макар и да не са изключително прецизни, са носители на ценна допълнителна информация за потребителските посещения, за интересите на потребителите, за тяхното поведение. Те имат голямо значение за персонализирането на услугите и за насочването на потребителите към информация, която би била максимално полезна за тях. Могат да са и в помощ на маркетинговите специалисти за изучаване и сегментиране на пазара. Освен това, този вид анализи се използват и за изучаване на конкуренцията, което е важен фактор за успешен бизнес.

Извличането на знания от Интернет ресурси може да бъде полезно в дейностите по управление на връзките с клиентите в следните основни насоки:

- **идентификация и групиране на клиентите.** На основата на обработка на данните, извлечени предимно от лог файловете (текстови файлове, наричани още журнали или дневници където се записват посещенията на уеб страниците) може да се осъществи идентификация на клиентски групи за маркетингови цели. За тази цел се извършва обработка чрез алгоритмите за клъстеризация, които позволяват търсене на независими групи и техните характеристики в набор от данни. Групирането на клиентите е предпоставка за сегментирането им и прилагането на диференцирани подходи при обслужването им, защото се знае се, че потребителите имат различни потребности и изисквания към стоките и харесват различни начини за покупка.

- **оценяване на клиентите.** Идентифицирането на най-ценни и лоялни клиенти може да се направи въз основа на данните от посещенията на уеб страниците и чрез анализ на записаните мнения във форумите. Класификацията и клъстеризацията на клиентските посещения показва какви са интересите на новите и съществуващите клиенти. Обработката на мненията на клиентите чрез методите за извличане на знания от съдържанието на текстови документи спомага да се идентифицират тези клиенти, които по-често записват положителни отзиви за стоките и съответно да се класифицират към групата на ценните клиенти, към които могат да се съставят персонализирани оферти.

- **подпомагане на персоналното обслужване.** Извличането на знания за предпочитанията и нагласите на клиентите към отделни продукти, услуги дава възможност за индивидуален подход при обслужването на всеки потребител. Следенето на предпочитанията и нуждите на клиентите позволява прилагането на специално отношение към всеки. Целият набор от дейности, свързани с грижата и вниманието към клиентите гарантира тяхното задържане. Анализите на посещаемостта на уеб страниците чрез обработка на сървърните лог файлове може да са източник на знания и за потребителските предпочитания, например разглежданите страници помагат да се натрупа информация за това какво харесва съответният потребител и ако например той си е закупил фотоапарат и разглежда аксесоари за него може да му се отправи специална оферта и той да се подтикне този към допълнителни покупки.

- **препоръки за нови покупки.** Процесът на генериране на препоръки е важен, защото клиенти, към които се насочват подходящи и навременни фирмени препоръки, купуват повече и са по-лоялни. Добър пример в това отношение е онлайн магазинът Amazon.com, при които се извеждат множество препоръки от типа “Потребителите закупили тази стока са закупили и ...”,

“Посетителите, които са разгледали тази стока са се заинтересували и от ...”, “Специално за вас...”. За всички тези препоръки са необходими не само данните за покупките, но и извлечените данни от съдържанието на страниците и от лог файловете, отразяващи посещаемостта на уеб сайта.

- **определяне на точния момент за предлагането на оферти.** На базата на данните от клик потока върху уеб страниците и анализ на поведението на потребителите може да се направят изводи за това кога е най-добре да се предлагат промоционалните оферти на клиентите. Например в момента, в който клиентите появяват интерес към дадена стока да им се предлагат промоционални оферти за нея или за свързани с нея стоки.

- **откриване на последователности.** Обработката на извлечените данни от Интернет източници чрез методите за асоцииране и откриване на събития, спомагат за откриване на събития, които се случват заедно и стоки, които се купуват заедно и могат да бъдат източник за намиране на закономерности в поведението на потребителите. Анализът на последователностите помага за увеличаване на удовлетвореността на клиентите, например има ситуации, при които между някои стоки няма видима връзка, но анализът на данните за направените покупки и идентифицираните разглеждани страници показват, че често клиентите се интересуват от дадена комбинация от стоки и е целесъобразно им да се предложи промоционална оферта за закупуване на стоките в пакет.

- **определяне на точните групи клиенти за дадена рекламна кампания.** Чрез предварителен анализ на данните от потребителските форуми и мнения, изследване на рейтинга на страниците и данните за посещаемостта на уеб сайтовете може да се стесни обхватът на групите клиенти, които е най-вероятно да откликнат на рекламите. Практиката е доказала, че в много от случаите повече нови клиенти се печелят чрез тясно фокусирани маркетингови кампании.

За да се открие по-добре на базата на кои видове Интернет източници може да се извличат знания и правят анализи за предложените насоки в областта на управлението на връзките с клиентите сме разработили табл. 1.

Таблица 1

Приложение на WM анализите в CRM

CRM дейности, в които са приложими технологиите за извличане на знания от Интернет ресурси	Извличане на знания от Интернет ресурси		
	Извличане на знания от съдържанието на Интернет източниците	Извличане на знания от структурата на Интернет източниците	Извличане на знания от данните за използваемостта на Интернет източниците
Идентификация и групиране на клиентите	•	•	•
Оценяване на клиентите	•		•
Подпомагане на персоналното обслужване	•	•	•
Препоръки за нови покупки	•	•	•
Определяне на точния момент за предлагането на оферти	•	•	•
Определяне на точните групи клиенти за дадена рекламна кампания	•	•	•
Откриване на последователности	•		•

Извличането на нови знания за клиентите се осъществява най-вече на базата на записаните данни в резултат от взаимодействието на потребителите с уеб сайтовете и чрез обработка на налична текстова информация, която те оставят под формата на коментари, съобщения във форуми и в писмата от e-mail кореспонденцията. Извлеченото ново знание

допълва възможност за осъществяване на по-точни анализи на покупателното поведение на клиентите и на факторите, които оказват влияние при вземането на решение за покупка. Освен това получените допълнителни знания за клиентите позволяват на фирмите да прилагат редица гъвкави инструменти и средства на маркетинга.

От технологична гледна точка за осъществяването на извличане на знания от Интернет ресурси не съществува един универсален алгоритъм, тъй като се работи с твърде разнообразни източници на данни. Могат да се използват софтуерни средства, които са предназначени за работа само с конкретен вид Интернет ресурси или такива, които са с по-общо предназначение и работят с повече видове данни. Трябва да се отбележи, че изборът на конкретно софтуерно решение зависи от множество фактори, като функционални възможности, възможности за интеграция със съществуващия CRM софтуер, цена. Няма универсално и най-добро решение за извличане на знания от Интернет ресурси, а трябва да се изследва приложимостта на всеки отделен пакет за конкретния вид бизнес и организация. Необходимо е да се отбележи, че при WM анализите съществено значение има човешкият фактор - веб анализаторите, от които очаква да определят основните параметри на анализите, да прилагат иновативни подходи и да интегрират избраните инструменти във фирмената аналитична информационна система. Според изследванията само 10% бюджета се изразходват за инструменти за анализ, а 90% за човешки ресурси, които да осъществяват процеса<sup>6</sup>.

В заключение може да се обобщи, че WM анализите макар и сложни за осъществяване могат успешно да се използват за подпомагане на дейностите по управление на връзките с клиентите. Чрез тях се получава допълнителна полезна информация за клиентите и се създават условия за прилагане на индивидуален подход при съпровождането на всеки клиент, както и възможности за спечелването на лоялността на клиентите и тяхното дългосрочно задържане.

#### **Използвана литература:**

1. Барсегян, А. и др. Анализ данных и процессов. 3-е издание. Изд. “БХВ-Петербург”, Санкт-Петербург, 2009
2. Markov, Z. and D. Larosed. Data Mining the Web Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons, New Jersey, 2007
3. Internet World Stats, <http://www.internetworldstats.com/stats.htm> (30.09.2015)
4. Кръстева, Н. Законът на Мур е в сила и за Интернет ресурсите, [http://cio.bg/4380\\_zakonat\\_na\\_mur\\_e\\_v\\_sila\\_i\\_za\\_internet\\_resursite](http://cio.bg/4380_zakonat_na_mur_e_v_sila_i_za_internet_resursite), (30.09.2015)
5. Etzioni, O. The World Wide Web: quagmire or gold mine?, Communications of the ACM, no. 11/1996, p. 65-68
6. Cooley, R. Mobasher, B. and Srivastave, J., Web Mining: Information and Pattern Discovery on the World Wide Web, <http://maya.cs.depaul.edu/classes/ect584/papers/cms-tai.pdf>, (9.12.2015).
7. Kaushik, A. Web Analytics. An Hour a Day, Wiley Publishing, 2007
8. Markov, Z. and D. Larosed. Data Mining the Web Uncovering Patterns in Web Content, Structure, and Usage. John Wiley & Sons, New Jersey, 2007
9. Kosala, R. and H. Blockeel. Web Mining Research: A Survey. <http://facweb.cs.depaul.edu/mobasher/classes/ect584/papers/kosala.pdf>, (1.10.2015).
10. Cronin, B. Annual review of information science and technology, Vol. 38, Information today Inc., 2004

#### **За контакти:**

Доц. д-р Снежана Динева Сълова  
Икономически университет – Варна  
E-mail: [ssulova@ue-varna.bg](mailto:ssulova@ue-varna.bg)

<sup>6</sup> Kaushik, A. Web Analytics. An Hour a Day, Wiley Publishing, 2007, p. 81.