

# Spotting The ‘Elusive’ Prospect Customer: Exploratory Study Of A Web-Powered Customer Relationship Management Framework

Myriam Ertz, Université du Québec À Montréal, Canada

Raoul Graf, Université du Québec À Montréal, Canada

## ABSTRACT

*Research on how Web-Mining (WM) optimizes marketing, is sparse. Especially absent, is research on WM usefulness for Customer Relationship Management (CRM). The purpose of this research, is to propose a Web Mining-enabled knowledge acquisition framework for analytical CRM. An exploratory study consisting of eleven in-depth interviews with marketing scholars and practitioners revealed that, WM methods and techniques - currently available to practitioners - are well-suited for identifying the profile of web prospects according to their browsing behaviour and to classify them into homogeneous groups. Besides, the nascent technologies regarding opinion mining, sentiment analysis or natural language parsing, and which underlie WM, seem sufficient to acquire knowledge pertaining to attitudinal and other more psychometrically-based characteristics about web prospects. Such tools enable to better understand the so-often termed ‘elusive’ prospects, by crafting fine-grained online marketing strategies to acquire those would-be customers. The authors discuss the managerial implications that derive from these findings.*

**Keywords:** Web-Mining; Analytic Customer Relationship Management; Prospect Consumer; Information Systems

## INTRODUCTION

The confluence of advances in Information Systems (Ranjan & Bhatnagar, 2011), rise of e-commerce (Bower & Maxham, 2012) and the relational paradigm of marketing (Grönroos, 1994) have enabled granular relationships with organizations’ client base. Relevant and meaningful customer information lies at the very core of this entangled process. Shaw, Subramaniam, Tan and Welge (2001), emphasized, more than one decade ago, that marketing depends more and more on the web for customer data. This trend is currently exacerbated by the inflating flow of data emanating from social networks, wikis, and other forms of User Generated Content, among others.

A powerful way to extract useful information out of gross web data is Web-Mining (WM) or the application of Data-Mining (DM) techniques to web data. Currently, most companies, not only own a website, but are also active on social media which provides them with a continuous stream of highly specific data. WM may therefore be very useful in order to obtain useful information about ways to improve relationships with customers. Its inductive approach and observational characteristics also overcome the hurdles usually associated with traditional survey-based marketing research *e.g.* low response rates, biases and errors (Mihai, 2009).

It has already been investigated to what extent WM provides interesting venues to profiling the behaviour of existing web customers (Ertz & Graf, 2013). However, little has been investigated regarding the potential of WM for identifying prospective customers online. More specifically, the literature about the integration of WM into broader Information systems *e.g.* Customer Relationship Management for processing gross web data about web prospects into useful information about these prospects is relatively scarce.

The first research question is thus, to what extent do WM methods applied to web data segment prospective web customers of a website? The second research question is, to what extent do WM methods applied to web data enable to collect useful knowledge about prospects' characteristics in order to craft targeted e-marketing strategies and acquire these prospects? First, a literature review provides theoretical conceptualizations and underpinnings for WM and aCRM and the subsequent research proposition. Second, the methodology and the profile of informants are presented. Third, results are provided and discussed. Fourth, implications for marketing practitioners are examined.

## **LITERATURE REVIEW**

### **The Definition of Web-Mining (WM)**

The WM research area converges from several research communities, such as Database (DB), Information Retrieval (IR), Artificial Intelligence (AI), Machine Learning (ML), Natural Language Parsing (NLP) (Kosala & Blockeel, 2000), Opinion Mining (OM) and Sentiment Analysis (SA) (Liu, 2011). WM draws actually on Data-Mining (DM) and related techniques that are used to automatically discover and extract information from web documents and services (Cooley, Mobasher & Srivastava, 1997; Kosala & Blockeel, 2000; Srivastava, Cooley, Deshpande & Tan, 2000).

Input formats of WM consist typically of intra-site and inter-site traffic data such as log files for global analyses (many users' browsing information on one or more websites), cookies for individual analysis (one user's browsing behaviour on one or more website) and personal identification information for personal analysis (one user's browsing behaviour on one specific site); transactional data (*e.g.* purchases); Information Systems data which may come from offline sources (*e.g.* Enterprise Resource Planning Systems [ERPS]); or various other content data such as emails, registration data, images or videos.

These formats support data from different nature such as Recency Monetary Frequency (RFM)-values of a users' visits, her profitability or economic value (difference between profits from a customer, segment or market and all the costs incurred to generate those profits), her lifetime values (updated net value - including all the "predictives" such as propensity, attrition or risk - of the expected future financial transactions with that user), data on products and contracts (numbers, mean product life), channel (preferred channel for ordering), relational data (interactions types with the online business) attitudinal data (customer loyalty patterns), psychographic data (lifestyle, personality, values, risk aversion, knowledge, focus of interest, opinions and behaviour, social styles), technical data (type of customer, payer status), sociodemographic or geographic data (Tufféry, 2011).

Outputs of WM consist of reports, segments, ready-to-use models, intelligent rules or direct recommendations for actions (Chiu & Tavella, 2008). They can be used for re-designing web site(s) (reorganization) or enriching the database(s) with models to generate personalized contents. Enriching the database requires to perform regular batch analyses. However, integrating the model directly to the website enables to apply the model in a real-time fashion to any new web user which saves time and costs (Bazsalicza & Naim, 2001). At the end of the 1990S already, Büchner, Anand, Mulvenna & Hughes (1999), noticed that WM has become a critical element of the marketing research process because it provides useful market(ing) intelligence.

Pattern discovery lies at the core of WM (Srivastava, Cooley, Deshpande & Tan, 2000). In this study, the stress will be put on the most important and common pattern discovery and analysis methods and techniques of WM. Others may exist but for conciseness and practical purposes, the focus has been kept on the major ones.

### **Web-Mining Methods and Techniques**

WM methods are similar to DM methods. Actually, in WM, DM methods and techniques are merely applied to the web data, which is also why WM is often called 'Web Data Mining'. For the sake of this research, these methods will be referred to WM methods. These are traditionally divided into four main categories summarized in Table 1.

*Association analysis.* Association Analysis is mainly used to find all co-occurrence relationships called ‘associations’, among data items (Liu, 2007). It thus analyzes the pages visited by a web user as well as the products bought by a customer on a single visit online (Tufféry, 2011). The classic application of association rule mining is the ‘market basket analysis’ (Agrawal, Imielinski & Swami, 1993). In association analysis, the challenge is to process enormous volumes of data and to identify those that are often happening together (Liu, 2007). For example, one may identify on an online book retailer website, that consumers who bought a book about DM also purchased a book about WM. This finding is an association that points toward specific cross-selling opportunities through the use of recommendation agents (e.g. Amazon’ “those who purchased this book X also purchased Y”).

*Clustering.* Clustering refers to segmentation and is therefore the division of a population of customers into a certain number of homogeneous subgroups, in order to enable organizations to better adapt their marketing strategies to each of these subgroups (Lendrevie & Lindon, 1993). Both clustering and association analysis are descriptive methods called ‘unsupervised learning approaches’.

*Classification.* Classification is one of the two major predictive and supervised learning methods of DM. Classifying behaviour or attitudes is particularly valuable in marketing. It enables, for example, to determine the probability that a prospect might become a customer, whether a customer is likely to leave or whether a customer may not be solvable. Scoring applications, such as attrition scores, epitomize that method. These applications identify whether a customer might renew a given contract subscription or continue buying from the company. The class variable (customer or non-customer) is determined by other various variables (e.g. demographics), which may predict the class variable. Then cases are gathered where class variable and the other pre-defined variables are available. A model is then built in order to link the class variable and the explaining variables. This model is eventually tested before being widespread and used for other applications (Bazsalicza & Naim, 2001). Hence, classification deals with discrete variables (e.g. client versus non-client).

*Prediction.* Prediction deals specifically with continuous variables (e.g. the potential amount of money spent on a purchase). This second major supervised learning method is complementary to the classification method. It implies to identify the class variable (amount of purchase), and other variables, which may explain the class variable. As for the classification method, a model is built to link these two types of variables and the model is verified before being generalized.

Each task entails a number of specific techniques that are applicable depending on the type of project undertaken and the outputs that are sought. Bazsalicza and Naim (2001) underline the importance of taking into account the whole scope and context in which the WM project takes place in order to determine the most adequate technique to be used depending on data types, outputs sought. DM and hence WM is always a process linked to a specific organizational objective. Table 1 summarizes the four main methods of WM.

**Table 1.** Categorization of Web-Mining methods

METHODS	GOAL	EXAMPLE
<b>DESCRIPTIVE MODELING</b>		
Clustering	Creating homogeneous groups	Segmenting consumers on two variables: “the amount of time spent on the website” and “the number of items downloaded”
Association	Finding sets of data items that occur together frequently (association rules) or that occur together frequently in some sequences (sequential patterns)	Finding regularities in the web data: in WUM, association rule mining can be used to find users’ visit and purchase patterns. Sequential patterns can be used to find users’ navigation patterns by analyzing clickstreams in server logs
<b>PREDICTIVE MODELING</b>		
Classification	Explain or predict the qualitative characteristic of an individual based on other qualitative characteristics of that individual	In the banking context, connecting socio-demographic characteristics of a customer to his/her possession of a specific financial product
Prediction	Explain or predict the quantitative characteristic of an individual based on other quantitative characteristics of that individual	In the telecommunication context, connecting consumption of customers to their quantitative socio-demographics

**Analytic Customer Relationship Management (aCRM)**

The customer is a strategic element in a company’s downstream supply chain (Xu & Walton, 2005). The changes of attitudes, behaviour, preferences of customers need to be identified by businesses in order to remain profitable. As Drucker (1996), put it: knowledge is the only meaningful resource. Gaining this knowledge is becoming an important differentiator for competitive advantage (Paiva, Roth & Fensterseifer, 2002). This is why managing customer relationships, is of utmost importance for market-oriented companies. CRM is a process designed to collect data related to customers, to grasp features of customers and to apply those qualities in specific marketing activities (Swift, 2001).

Theoretically, CRM leverages and exploits interactions with customers to maximize customer satisfaction, ensure business returns and enhance customer profitability (Xu & Walton, 2005). More specifically, it is a combination of business processes and technologies in order to understand a company’s customer from different perspectives in order to differentiate products and services of that company competitively (Tiwana, 2001).

However, in practice, it appears that CRM has become more of a buzzword (Luck & Lancaster, 2003). There appears to be a fundamental problem in CRM research at present because no common image regarding what CRM is actually exists (Choy, Fan & Lo, 2003). It appears to mean different things to different people (Paulissen, Mills, Brengman, Fjermestad & Romano, 2007). In fact, CRM is difficult to conceptualize because it is a cross-disciplinary field of research which includes marketing, business management, IT and Information Systems (Ngai, 2005). The CRM concept is also relatively new since it appeared at the end of the 1990s (Romano & Fjermestad, 2002a,b) and has kept on increasing in importance in academia since then, until 2004 (Wahlberg, Strandberg, Sundberg & Sandberg, 2009).

Wahlberg, Strandberg, Sundberg and Sandberg (2009) argue that there are four different perspectives applied to the phenomenon of CRM in the literature: (1) CRM being a matter of integrating business processes in an organization; (2) CRM being a matter of customer-focused business strategy; (3) CRM being a matter of customer knowledge management; and (4) CRM being a matter of technology-enabled customer information management activities, including Strategic CRM (sCRM), Analytical CRM (aCRM), Operational CRM (oCRM), Collaborative CRM (cCRM), Technical CRM (tCRM) and according to some authors Electronic CRM (e-CRM) (Xu & Walton, 2005; Romano & Fjermestad, 2002a,b; Bazsalicza & Naim, 2001). The latter perspective (number 4) subdivides CRM into relevant categories, including an analytical one, which suits well the purpose of this research. Therefore, it will be used throughout this piece of research, by focusing namely on Analytical CRM (aCRM).

The aCRM branch develops a customer knowledge database from systematic collecting and storing of customer data (Wahlberg, Strandberg, Sundberg & Sandberg, 2009). It provides a 360° view of the customer

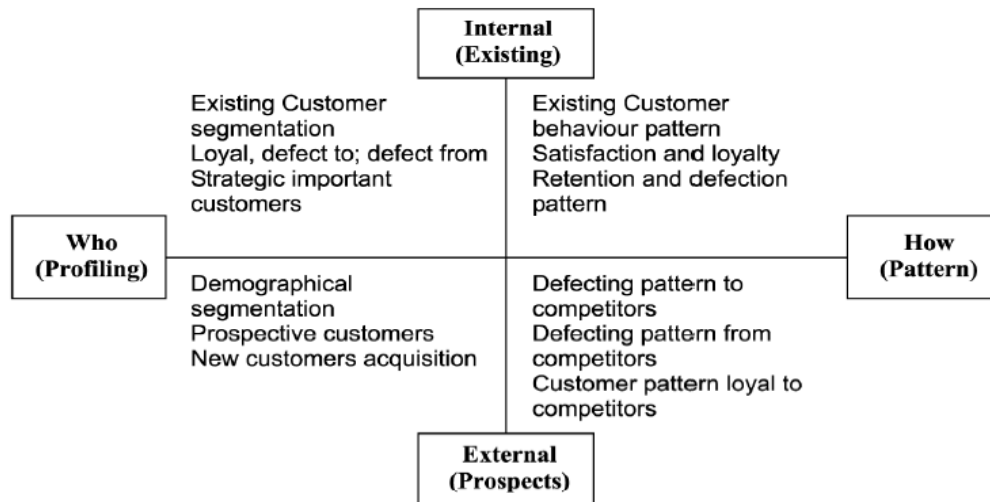
(Kotorov, 2002). The field of aCRM is dominated by the DM concept which collects primarily offline “islands of customers’ data” throughout the organization, with data warehousing techniques (Wahlberg, Strandberg, Sundberg & Sandberg, 2009).

Online web data (Hyperlinks, web log files, texts, images, and so forth) complement those “islands of offline data”, by completing them with additional information on customers and prospects, obtained from the internet. The offline-based aCRM process encompasses powerful predictive techniques such as DM, forecasting and scoring (Kimball & Ross, 2002). They typically segment customers more effectively or optimize offerings to better fit customers’ buying profiles (Xu & alton, 2005). Adding web data to the aCRM is thus expected to optimize that process since it will leverage more powerful marketing capacities.

Analytical DM tools, may enable to discover non-obvious or non-linear patterns in the data, in order to generate customer profiles and support segmentation efforts in marketing and promotion strategies. Such techniques may thus also be worth being applied on web data emanating from web prospects, on top of those generated by existing web customer, in order to increase the customer base by attracting additional would-be customers.

According to Ranjan and Bhatnagar (2011), the aCRM objectives that are relevant to marketing are: (1) Helping in cross-selling and up-selling; (2) Allowing targeted marketing; (3) Helping in segmenting customers on the basis of fixed criteria. Xu and Walton (2005) developed a more exhaustive framework including the tasks of aCRM according to the “who” or profiling tasks and the “how” or the patterns identification tasks, of simultaneously existing customers (internal) and prospect customers (external). The framework for customer knowledge acquisition is presented in figure 1.

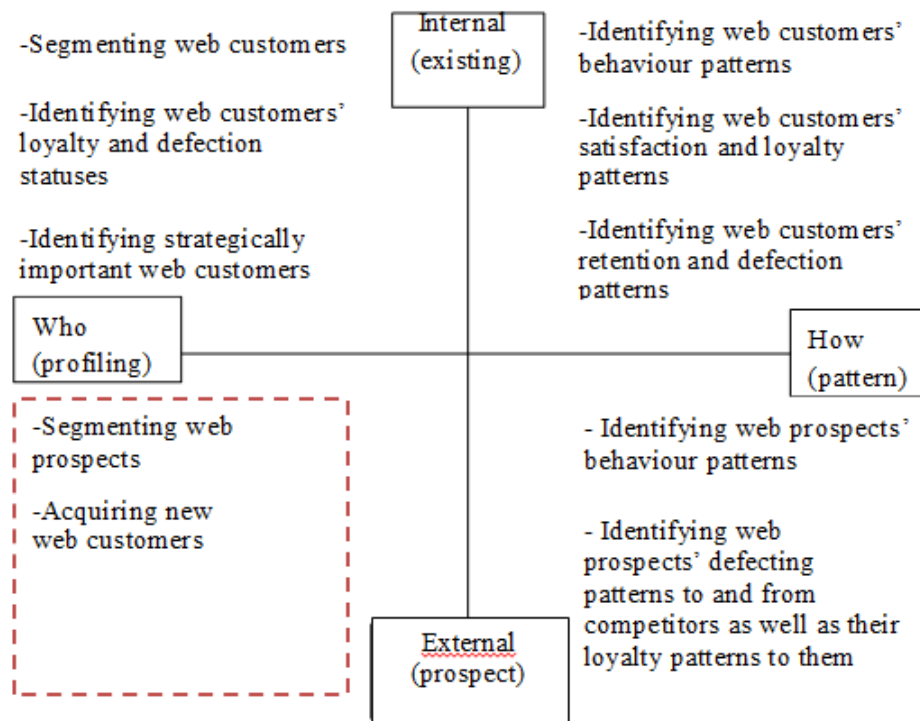
**Figure 1.** Xu and Walton’s (2005) aCRM for a customer knowledge acquisition framework



**A Framework of aCRM for Web Users’ Knowledge Acquisition**

Xu and Walton’s (2005) framework refers primarily to an offline-based aCRM process. By adapting their model, it is, however, possible to develop a framework of aCRM for web users’ knowledge acquisition, as illustrated in Figure 2 below. WM methods used on web data enable to achieve the main aCRM objectives which can be placed in a two- dimension conceptual map: the “prospects” axis (ordinate) ranges from prospects to existing customers, while the “research type” axis (abscissa) ranges from profiling study (who are the users?) to behaviour study (how do users behave?). The use of WM methods and techniques is assumed to turn operational web data into meaningful and relevant knowledge of current and prospective web customers. Integrating the WM process into the aCRM process provides a formal and structured guideline which improves efficiency and effectiveness of the WM process while providing insightful knowledge to the aCRM applet of CRM. This study focuses on the two objectives situated in the lower-left quadrant, framed in red, dotted lines.

Figure 2. aCRM for a web users’ knowledge acquisition framework



Prospective customers of a website are defined as those who visit the website without having ever made a purchase on it. Any other definition of a prospect would be difficult to make in a web context. Consequently, all visitors are considered prospects as long as they have not made a purchase, or performed the desired transactional action. If they do, they are converted into existing customers who were the focus of the two previous sections.

There are many advantages related to customer profiling (Van Wel & Royakkers, 2004). It is often more cost efficient to look at a group of web users instead of looking at each individual, because groups are cheaper and easier to approach (for instance by placing an ad in the right website or web page instead of mailing every individual member of the group) (Custers, 2001). These so-called group profiles can also be of added value to individual profiles, because of the fact that some individual characteristics only become clear after looking at the individual from a group perspective (Custers, 2001). Clustering web users can thus also be used for marketing strategies, personalization and web site adaptation (Song & Shepperd, 2006). The characteristics of web usage can be used as a basis for clustering and segmentation (Srivastava, Cooley, Deshpande & Tan, 2000). Web browsing patterns can also be detected from a group point of view with web user clustering (Song & Shepperd, 2006). Such patterns can complement web users’ profiles that had already been developed beforehand.

**RESEARCH PROPOSTIONS**

Segmentation of prospects is very similar to that of existing customers. The main difference lies in the lack of prospects’ transactional data. However, prospects leave traces when they visit a website (e.g. form fill-ins, log files, caches, cookies), which provide hints as to their personal and private characteristics. By using such data in an aggregated and anonymous fashion, web businesses can create highly detailed sociodemographics-based profiles of web prospects (Tuffery, 2007), at a specific point in time. Such surgical profiling optimized in turn segmentation and therefore mass customization or morphing capacities. Hence, the following Research Propositions (RP) are expected.

**RP 1:** web data generated by prospective web customers are sufficiently detailed and accurate to provide a strong basis for the creation of precise profiles about prospective web customers.

**RP 2:** clustering, association analysis, classification and prediction methods applied to web data create homogeneous groups of prospective web customers.

In their original framework, Xu and Walton (2005) based segmentation on demographic data. In order to develop e-marketing strategies and especially direct marketing strategies, as seen before, web managers need to have knowledge of their prospects' profiles which may be comprised of wider sorts of characteristics than sociodemographics alone. Useful knowledge, in that respect, refers to prospects' preferences, needs and wants. If profiling does not take into account the wide array of prospects' interests and if segmentation is not based on such attributes, drawing relevant e-marketing strategies to acquire prospects might not be flawless.

The main benefit of transactional data is to provide insight into the needs, wants, interests or preferences of existing customers (Tufféry, 2011). Such knowledge is thus not accessible for prospects since they are not customers of the website (yet). It is therefore usually inferred from other sources such as browsing behaviour or clickstream data (Mihai, 2009). Currently, social media information and data, have created plenty of additional opportunities to buy and aggregate highly detailed knowledge about almost every aspect of a human's life (e.g. professional life with LinkedIn and Twitter, private life with Facebook, leisure/hobbies with Flickr, YouTube, Pinterest or Tumblr) (Wilcox & Stephen, 2013; Naylor, Lamberton & West, 2012). Segmentation may thus be completed on a multitude of attributes such as geographic data (e.g. ZIP code); psychographic data (e.g. values, personality, lifestyle, preferences, opinions); or behavioural variables (e.g. purchase frequency, time spent during a website visit) (Tufféry, 2011; Rui, Dholakia, Chen & Algesheimer, 2012).

WM offers tremendous capabilities regarding the mining of this ever-growing flow of data. With WM tools, social media, (micro-)blogs, webpages or personal homepages can be collected and after interpreting the unstructured text information on these formats into business intelligence stored in a database, they can be used for marketing purposes (Lau, Lee & Ho, 2005). Sentiment Analysis and Opinion Mining, in particular allow to understanding time-related aspects of consumer behaviour as well as their evolution through time (Liu, 2011). Opinions, for example, can be extracted from various web sources containing structured or unstructured text, such as customer written comments, blog posts or written feedback, in order to perform opinion mining (Liu, Zhao, Benkler & Xiao, 2006). The web content to be mined may also consist of images, audio or video data (Yeh, Lien, Ting & Liu, 2009). The visitors' web usage patterns and trends for e-commerce might also be predicted (Abraham & Ramos, 2003). Extracting usage patterns enables organizations to determine the life-time value of customers, deploy marketing strategies across products and monitor the effectiveness of marketing campaigns (Cooley, Mobasher & Srivastava, 2003).

These numerous WM approaches enable to adapt the content and structure of the website according to a specific segment of users, be it on the basis of demographic, psychographic, geographic or browsing characteristics. It also enables mass customization leading to a more use-centered approach and may thus be considered as an indispensable part of e-commerce (Van Wel & Royackers, 2004). Businesses can then develop highly customized advertising tailored to the specific needs and interests of prospective customers at a specific point in time and diffuse them on all the web pages the prospects might visit. This display advertising technique known as behavioural targeting has proven to be highly effective at increasing publishers' (those who supply advertising space online) revenues through increased click-through rates by 670% (Yan *et al.*, 2009). As a result, there are higher conversion rates and increased revenues for web businesses as well (Beales, 2010). The following Research Propositions (RP) are therefore expected:

**RP 3:** web data generated by prospective web customers are sufficiently detailed and accurate to provide various psychometric characteristics (e.g. values, preferences, lifestyles) about those prospective web customers, to be used for acquiring those prospects.

**RP 4:** WM methods applied to web data generate relevant information about prospective web customers' characteristics which can be used for subsequent targeted marketing and sales efforts to acquire them.

**METHODOLOGY**

Lack of relevant scientific sources on WM applied to marketing called for an exploratory research design in order to explore WM experts’ opinions about current WM potentialities regarding identification of prospective web customers on the web. An open-ended research questionnaire investigated informants’ opinions on the capabilities of WM to fulfill the two objectives of aCRM related to the identification of web prospects’ behaviours online. A short lexicon defined the different WM methods to prevent misinterpretations. A convenience sample was drawn and the researcher was able to conduct eleven valid semi-structured interviews. The number could not be higher since only those informants who had a thorough knowledge of WM and a sound understanding of marketing issues, qualified for the interview, which is rather rare. A summary of the informants is provided in table 2. About half of them were C-level executives in public or private organizations while the other half consisted of academicians. Their area of expertise ranged from marketing, communication and IT to engineering and R&D. They were between 36-55 years old and one informant was more than 55 years old. The sample was heterogeneous enough to allow for a diversity of opinions and responses (Malhotra, 2010). Most in-depth interviews were conducted face-to-face and lasted between one and two hours. Answers were completely transcribed and summarized into a response matrix. A content analysis led to a synthesis of all answers to (in)validate research propositions, and ultimately, research questions.

**Table 2.** Profiles of informants

<b>Job Title</b>	<b>Company Sector</b>	<b>Area of Expertise</b>	<b>Age Range</b>
Marketing And Brand Development Director	Insurance	Marketing Communication	36 – 55
Marketing Research Agency Co- Owner	Marketing Research And Analytics	Marketing Communication	55+
Web Business Manager	Academia	Marketing Communication IT Other	36 - 55
Database Administrator (DBA)	It	IT	36 – 55
Scholar	Academia	Marketing Communication IT	36 – 55
Marketing Research Agency Director	Marketing Research & IT	IT Engineering R&D	36 – 55
Scholar	Academia	Marketing Communication	36 – 55
Scholar	Academia	Other: statisticsand data-mining	36 – 55
Scholar	Academia	Marketing Communication	N.S.
Scholar	Academia	Marketing Communication Other: statistics, scoring and Data-mining	N.S.
Scholar	Academia	Marketing Communication	N.S.

**RESULTS AND DISCUSSION**

**Segmenting Prospective Web Customers of a Website**

Out of the 11 informants, 9 indicated that WM methods were effective in identifying the profile of prospects. Eight informants agreed on the value of classification for segmentation purposes. Three of them added that when combined to regression analysis, WM offers tremendous insight into the interests of prospects which enables to segment them, whereas the fourth informant indicated that regression was not that useful in that respect. One respondent further explains:

*“This can be summarized as follows: show me how you click on the website and I will tell you who you are and what you may be likely to do on the website as well as your future behaviour.”*

Website managers own useful browsing information, clickstreams and socio-demographic as well as socioeconomic information collected from forms, registrations, etc. in order to draw useful profiles with classic WM tools. The level of segmentation determines the resulting marketing strategy from mass to individual. Social networks which provide a wide array of rich attributes on which to segment prospects or garner a given individual’s profile. Gaining useful insight of prospects, is a stepping-stone for segmenting them and crafting



strategies accordingly to acquire them in multichannel settings, through tailored display advertising (behavioral targeting, demographic targeting, geographic targeting) or direct marketing campaigns, the very core of CRM. One respondent puts it this way:

*“The potential market (prospects) is not only online or offline, it can and is usually both. It is important to have an overall, a global view of the market.”*

Both RP1 and RP2 are therefore validated. Consequently RQ 1 stipulating whether WM methods and techniques enable to segment prospective web customers, is answered.

### **Prospect Knowledge Collection for Targeted E-Marketing Strategies Development**

Prospects may be segmented against a multitude of attributes (Values and Lifestyles, Rokeach models on values and lifestyles, Attitude Interest Opinion model). Such data can be gleaned successfully in a web context with slight differences from real-world data in some instances. The rise of social networks has much helped in that respect because they are spots where people reveal their true preferences, habits and needs without even noticing it. One respondent highlighted that:

*“FB does not sell demographic data they mainly sell behavioral data : links on which users click, ads they watch, what they “like”, what their friends “like” and all activities put at their disposal for their own pleasure, while they are actually sharp measurement and evaluation tools of behaviours. These are then sold.”*

Browsing paths, traffic data inform about a prospect’s web behaviour and hence interests, habits, preferences and the like. Transactional data provide additional and sharper insight into that respect and can be correlated to browsing behaviour in order to triangulate both data blocks and pinpoint true interests and habits. One respondent puts it like this:

*“Traffic analysis, browsing paths of prospect and longitudinal statistics allow to crosstabulate prospects and existing customers to identify correlations. This approach permits to link actual transactions to the website traffic, in other words to develop a relation between browsing volume (generated by both prospects and customers) and final transaction data (generated by old or new customers only) to develop such statistics as: for each additional unique new 2000 prospects, 1 purchase is made.”*

WM tools and to a lesser extent descriptive statistics enable to discover these behavioural patterns which are hints to habits and interests.

It appears that WM tools can be subdivided into two approaches: one “slow- and-steady” approach that analyzes the data and produces business models, rules or algorithms in a post-hoc fashion and another “on the fly” approach that produces real-time, dynamic response frameworks (models, rules), that are directly applicable on the website. The latter approach seems more efficient, especially in the fast-moving web context. One respondent suggests:

*“The usefulness of traditional slow-and-steady methods that keep up with the flight of individuals in their demographics flocks will become as obsolete as the Pony Express is to communications. Now users become savvier and internet grows consequently in its ability to respond to user preferences with dynamic approach to cross-selling based on regression/classification and the use of clustering that responds on the day or the hour according to current trends, which is far more valuable than traditional approach.”*

Outputs generated by both approaches have effects in the short as well as in the long run. In the short term, better knowledge of prospects’ interests and other psychographical data enables businesses to build highly tailored e-marketing strategies (e.g. Morphing, Behavioral Targeting), increasing in turn Click-Through Rates (CTR), Click Through Interests (CTI) or the views divided by the clicks and View-Through Rates (VTR) or the views divided by the shows, the conversion rates and acquisition rates.

The Impact of those strategies need not be instantaneous but may also impact prospects in the longer run through branding and conditioning to increase brand awareness, recall, recognition, equity to increase the brands', products', organisation's share of prospects' minds and hopefully share of heart (Kotler & Keller, 2006). Prospects may therefore consider the website during future decision-making processes. One respondent indicated that it is more useful to do branding and conditioning on the longer run:

*“If you are sent many ads corresponding to your tastes, even though you might not have the budget now, you will have it one day when you will start working and then you might consider those offers again (recall) because you will already have a favourable idea of the brand due to these targeted actions. You can convert prospects into customers in the longer run too.”*

Their likelihood to buy is thus low at the moment but may increase substantially in the longer run when more favourable conditions or settings enable the web visitor to be more 'spendthrift'. Both RP3 and RP4 are validated. Consequently RQ 2 stipulating whether WM methods and techniques enable to identify prospective web customers' psychographic characteristics to acquire them, is answered.

### **Discussion of the Results**

Unlike existing customers, prospects are harder to profile because the company lacks fundamental data such as transaction information that usually bring the most insight into individuals' profiles. With the Web, however, prospects are not those 'elusive' beings anymore since their behaviour can be tracked and analyzed. However, the issue that remains refers to the identification of each unique user and the linking of relevant web browsing data or content to that user, with confidence. A triangulation of datasets such as traffic analysis, browsing behaviour, clickstream data, longitudinal statistics (descriptive statistics done over time), profiles data or user-generated content, can then used as input to alleviate that issue. They just need to be large, well retrieved, well-selected, well-cleansed and pre-processed.

The business usually transposes the segmentation level applied to existing customers to prospects. As previously identified, the segmentation continuum ranges from masses to the individual. The targeting level chosen determines the level of analysis for the WM project (e.g. study of groups or individuals) and resulting marketing strategies can be used to attract and retain prospects after WM outputs have revealed relevant knowledge In that respect. The segmentation process can be done through the slow-and-steady approach or the real-time approach, which is quicker and thus more appropriate in a fast-paced environment as the Internet.

Either way, it has been established that WM contributes to draw very accurate specific prospects' profiles. The classification technique is very useful for estimating the missing transactional aspects of prospects. By using existing customers' transactional information, predictive models can be developed in order to anticipate the likely action or attribute of a prospect. Prospects can also be assigned a score which classifies them into identified groups. WM also allows determining more psychometrically-oriented variables such as prospects' interests, values and preferences. Based on the derived knowledge, the business can build segmentation schemes (e.g., psychographics-based niche segmenting). The choice of the variables to be used as group descriptors, as well as segmentation techniques and other arbitrary considerations remain under the control of the manager.

Those outputs that WM enables to obtain are framed in bold. Prospects' profiles together with their identified needs, interests, wants, habits and preferences drive the business' eMarketing acquisition efforts. It is based on that knowledge that the company can craft relevant strategies that have high chances to be considered by prospects because they correspond to their tastes and needs.

The effect of these acquisition strategies may be twofold: in the short run and in the long run. Most businesses will seek to get the short run impact which affects mostly the bottom line. In fact, deriving accurate prospects' profile and offering them a value proposition tailored to their unique being, increases the likelihood that these prospects will engage in the activity desired by the company be it a video game account subscription, a book purchase, or whatever else. In web advertising, behavioural advertising capitalizes on WM techniques to enhance core strategic indexes such as Clickthrough Rates, Viewthrough Rates or Clickthrough

Interest and boosts conversion rates (Yan *et al.*, 2009; Fulgoni, 2008). More clicks on ads means also more revenues for publishers (owners of the website displaying ads) or any other entity that hosts ads, enabling them to also enhance the quality of their website(s) (Beales, 2010).

For e-business announcers increased click rates also increase the Return on Marketing Investment (ROMI) comprised of Cost Per Mille, effective Cost Per Mille, Cost per Thousand, Cost Per Impression, Cost Per Click/Pay Per Click or Cost per Action, depending on the pricing model chosen. Such a virtuous circle makes e-commerce more efficiently-run. This is in line with what Pfeiffer and Zinnhauer (2010) asserted, namely that online advertising is especially effective in later-funnel stages (order, repeat visits, repeat order), but should still be combined to classic media to support upper-funnel marketing-related activities (impressions, clicks, offers, calls-to-action). WM-driven eMarketing tactics and strategies may therefore have a tremendous impact on e-business' KPIs and market shares in the short term.

Meanwhile, additional long-lasting and incrementally-built impacts develop in the longer run. In the literature it has in fact been recognized that highly targeted online campaigns contribute to brand equity by enhancing brand status, namely brand recall, awareness, sympathy, intent to use and first choice (Pfeiffer & Zinnhauer, 2010; Hollis, 2005). Attitudinal aspect of brand equity (*i.e.* brand strength) can be estimated via the BrandDynamics framework (Dyson, Farr & Hollis, 1996). Other measurement tools may apply.

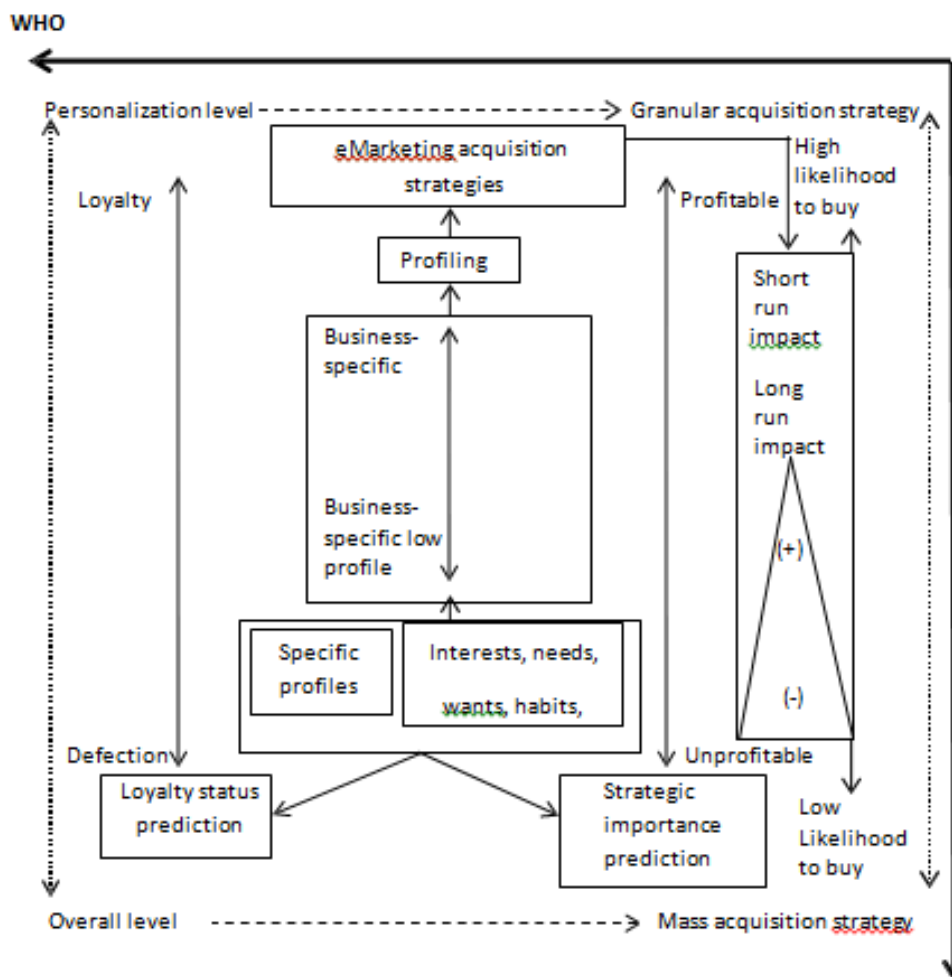
Online advertising contributes to increase the five levels of attitudinal loyalty to a brand to heighten the probability of a person buying a brand (Hollis, 2005). The lowest level is *presence* (knowledge of what the brand stands for). It corresponds to the lowest prospects' likelihood to buy; the *relevance* step is defined by negative drivers of loyalty (*e.g.* cheap brand thought to be of bad quality). In *performance*, the prospect agrees the brand provides acceptable levels of performance on basic criteria (Daniel's (1961) Critical Success Factors (CSFs)). In *advantage*, the prospect admits that the brand displays long-lasting and ongoing advantages over competitors (Daniel's (1961) Key Success Factors (KSFs)). Eventually, *bonding* takes account of prospects' relative classification of the brand's advantages (or KSFs) and the degree to which they believe the brand shares the same endorsement than others. By moving up the pyramid as shown in figure 3, the prospects' likelihood to buy increases accordingly. Both brand status and brand strength contributes to capture more of prospects' share of the mind and ultimately Share of the heart (Kotler & Keller, 2006).

The online marketing campaign remains framed in the segmentation level and subsequent acquisition strategy that have been chosen by the business. Mass segmentation and its subsequent global acquisition strategy will typically lead to less relevant and interesting marketing campaigns for prospects. This approach may nurture their knowledge of the brand but also decrease their purchase intentions. On the contrary, personalized segmentation and its subsequent granular acquisition strategy drives prospects to the bonding stage so that they agree by themselves that the brand is superior, different and worth the dollars to be spent.

The third theme corresponding to the third meta-objective of profiling prospective web customers of a website, as identified in Xu and Walton's (2005) adjusted framework, can be fulfilled by using WM. Traditional descriptive statistics as well as data about prospects issued by third parties or syndicated services constitute useful complementary tools and material in that respect.

Figure 3 depicts the potentialities of WM to reach the third aCRM meta-objective (theme) of profiling prospective web customers on the internet. WM is well-suited to fulfill the seventh objective of segmenting prospective web customers as well as the eight objective of discovering prospects' needs, wants, habits, preferences and interests, in Xu and Walton's (2005) web-adapted framework for aCRM knowledge acquisition. Depending on the segmentation level sought, web businesses can combine segmentation attributes, needs, to predict potential loyalty statuses of prospects on the business-specific loyalty continuum (loyal vs. defective) as well as their potential profitability on the business-specific profitability continuum (profitable vs. unprofitable). These attributes and knowledge on prospects enable businesses to draw business-specific profile (profiling) and craft profile-specific eMarketing acquisition strategies in accordance with the pre-defined segmentation level. These strategies, especially web advertising may have short term effects on the bottom-line or more long-term effects such as branding (brand status and strength) as well as increase in share of hearts or share of minds.

Figure 3. WM-enabled profiling of prospective web customers



**IMPLICATIONS**

This study fills in the gap in the marketing literature regarding the potentialities of WM for marketing. It represents a unique junction of two literature streams: the marketing literature and the information systems literature comprised of the specific CRM and WM components. It provides further theoretical underpinnings regarding the, overlooked, but tremendous potentialities of WM as integrated to the aCRM framework to yield insightful knowledge about prospective web customers who interact with online businesses e.g. e-commerce portals.

The exploratory nature of the research project provided a richer understanding of the benefits of WM methods and techniques on the analytical CRM (aCRM) applet of the marketing function, in the framework of a Knowledge Management (KM) perspective. The relationships identified in the model referring to the profiling of prospective web customers, can be tested empirically for external validity and generalizability.

In the light of this study, it is very important for managers and practitioners to understand that web-mining is not an end in itself. It is a combination of varied tools and techniques that enable experienced teams to discover useful meaning out of big (un)structured data. There is always a “garbage in-garbage out” pitfall, since all data is not necessarily good to process. Even all information derived from the data is not necessarily useful either. Data produced by logged in customers appear to be the best kind of data since navigation patterns

can be precisely attached to a specific customer. For prospects, it might therefore be wise to require them to register on the website and automatically log them in whenever they return on the website. Companies' databases should also allow for high volumes of data entries, since higher quantities of data leverage better results.

While WM is very well-suited to discover the profiles of prospective web customers of an e-commerce website, it should be stressed that WM can momentarily only process tangible, operational data which refer to browsing history, clickstreams, logging history, and so on. Consequently, it is an extremely powerful approach to analyze navigation patterns, web paths, uncover hidden or a priori non-obvious relationships in the web data. It informs thus well about the specific browsing behaviour and factual loyalty patterns development of both existing and prospective web customers. Segmentation might therefore be relatively well-done based on such characteristics. Besides, groupings and profiles based on more psychographic data (e.g. opinions, lifestyles, attitudes) might also be performed by means of WM. Recent advances in WM such as opinion mining, sentiment analyses and other semantics-related analyses have attempted to increase proficiency in exploring and discovering these more latent elements of web users' behaviours. However, these techniques are expert systems, very costly, calling for appropriate infrastructures and human resources, making their use still very limited. Future opportunities lie in democratizing access to such analytical tools and processes. Exploration and mining of social media data however provide rich promises in that respect since they refer to almost all characteristics, especially psychographic ones, of a prospect.

### CONCLUSION

This study sought to investigate the integration of WM methods and techniques to the analytical Customer Relationship Management (aCRM) applet is part of many Marketing Intelligence Systems. The specific conceptual framework started from an adapted version of Xu and Walton's (2005) customer knowledge acquisition framework, for web data. An exploratory study consisting of eleven in-depth, semi-structured interviews with experts in both WM and marketing revealed that WM is well-suited to achieve two objectives of the framework related to the profiling of prospective web customers: (1) to segment prospective web customers of a website; (2) to collect useful information about prospects to develop targeted e-marketing strategies to acquire these prospects. Besides, WM seems also appropriate in order to determine: (3) more psychographic variables about prospects such as their preferences, needs or habits and ; (4) the insight gained through WM is useful for crafting highly targeted marketing and sales efforts online. A thorough discussion of these results is provided. Moreover, managerial implications are also discussed.

### AUTHOR INFORMATION

**Myriam Ertz** is a PhD student in Marketing, Marketing Department, École des Sciences de la Gestion (ESG), Université du Québec à Montréal (UQÀM) [School of Management Sciences – University of Québec at Montréal]. Her research interests are related to online business and responsible consumer behaviour. Her research has appeared in journals such as the *International Journal of Electronic Commerce Studies* and conferences held by the European Marketing Academy (EMAC), the Administrative Sciences Association of Canada (ASAC), the Academy of Taiwan Information Sciences Research (ATISR) NETs conference, the Association Française du Marketing (AFM) and the Association Francophone pour le Savoir (ACFAS).

**Fabien Durif** is Professor of Marketing, Marketing Department, École des Sciences de la Gestion (ESG), Université du Québec à Montréal (UQÀM) [School of Management Sciences – University of Québec at Montréal]. His research interests lie in consumer behaviour, services marketing, relational marketing and marketing ethics. His research has appeared in such outlets as the *Journal of Business Research*, the *International Journal of Electronic Commerce and Studies* or the *Journal of Research in Interactive Marketing*, and proceedings of conferences held by the European Marketing Academy (EMAC) or the European Academy of Management (EURAM).

## ACKNOWLEDGEMENTS

This paper is based on a conference communication at the Conference of the Administrative Sciences Association of Canada (ASAC), at Huntsville, Canada, in May 2014. We are grateful to all the session participants for their valuable comments regarding improvements for this paper.

## REFERENCES

- Abraham, A. & Ramos, V. (2003). Web Usage Mining Using Artificial Ant Colony Clustering and Linear Genetic Programming. In *Evolutionary Computation*, pp.1384-1391.
- Agrawal, R., Imielinski, T. & Swami, A.N. (1993). Mining Association Rules Between Sets of Items in Large Databases. In Proc. of the *ACM SIGMOD Intl. Conf. on Management of Data (SIGMOD'93)*, pp.207-216.
- Bazsalicza, M. & Naim, P. (2001). *Data Mining pour le Web: Profiling, Filtrage Collaboratif, Personnalisation Client*. Eyrolles.
- Beales, H. (2010). The Value of Behavioral Targetin. Network Advertising Initiative (NAI): <<http://www.networkadvertising.org/index.asp>> (last visited Mar. 3, 2012)
- Bower, A & Maxham, J (2012). Return Shipping Policies of Online Retailers: Normative Assumptions and the Long-Term Consequences of Fee and Free Returns. *Journal of Marketing*, 76(5), pp.110-124.
- Büchner, A.G., Anand, S.S. Mulvenna, M.D. & Hughes, J.G. (1999). Discovering Internet Marketing Intelligence through Web Log Mining. In Proc. of the *Unicom99 Data Mining & Datawarehousing: Realising the full Value of Business Data*, pp.127-138.
- Chiu; S. & Tavella, D. (2008). *Data Mining and Market Intelligence for Optimal Marketing Returns*. Butterworth-Heinemann.
- Choy, K.L., Fan, K.K. & Lo, V. (2003). Development of an Intelligent Customer- Supplier Relationship Management System: the Application of Case-based Reasoning. *Industrial Management & Data Systems*, 103 (4), pp.263-274.
- Cooley R., Mobasher B. & Srivastava J., (2003). Data Preparation for Mining World Wide Web Browsing Patterns. *Knowledge and Information Systems*, 1 (1), pp. 5-32.
- Cooley, R., Mobasher, B & Srivastava, J. (1997). Web-Mining: Information and Pattern Discovery on the World Wide Web. In Proc. of the *ICTAI 1997*.
- Custers, B. (2001). Data mining and group profiling on the Internet. In *Ethics and the Internet*, A. Vedder (ed.), Antwerpen, Groningen, Oxford: Intersentia, pp.87-104.
- Daniel, R. (1961). Management Information Crisis. *Harvard Business Review*, Sept.- Oct.
- Dey, L. & Mirajul Haque, S.K. (2008). Opinion Mining from Noisy Text Data. In Proc. of the *Second Workshop on Analytics for Noisy Unstructured Text Data*.
- Drucker, P. (1996). The Information Executives Truly Need. *Harvard Business Review*, January-February, pp.54-62.
- Dyson P., Farr A. & Hollis N. (1996). Understanding, Measuring and Using Brand Equity. *Journal of Advertising Research*, 36 (6), pp.9-21.
- Ertz, M. & Graf, R. (2013). Benefits of Web-Mining For Profiling Existing Web Customers. In Proc. of the *Clute Institute International Academic Conference*, Key West, Florida, 380-382.
- Fulgoni, G.M. (2008). *How Online Advertizing Works Whither the Click,* ComScore, prepared for Empirical Generalizations in Advertising Conference for Industry and Academia, The Wharton School, Philadelphia, PA.
- Grönroos, C. (1994). From scientific management to service management: a management perspective for the age of service competition. *International Journal of Service Industry Management*. 5 (1), pp. 5-20.
- Hollis, N. (2005). Ten Years of Learning How Online Advertising Builds Brands. *Journal of Marketing Research*, June, pp.255-268.
- Kimball, R. & Ross, M. (2002). *The Data Warehouse Toolkit: The Complete Guide to Dimensional Modeling, 2<sup>nd</sup> Edition*. Wiley.
- Kosala, R. & Blockeel, H. (2000). Web Mining Research: A Survey. In Proc. of the *ACM SIGKDD Explorations*, Vol.2, pp.1-15.
- Kotler, P. & Keller, K.L. (2006). *Marketing Management, 12<sup>th</sup> Edition*.

- Kotorov, R. (2002). Ubiquitous Organisation: Organisational Design for e-CRM. *Business Process Management Journal*, 8 (3), pp.218-232.
- Lau K-N, Lee K.-H., Ho Y. (2005). Text Mining for the Hotel Industry. *Cornell Hotel and Restaurant Administration Quarterly*, 46 (3), pp. 344-362.
- Lendrevie, J. & Lindon, D. (1993). *Mercator – Théorie et Pratique du Marketing*. Editions Dalloz.
- Liu, B. (2007), *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data*. Springer. Verlag Press.
- Liu, B. (2011). *Web Data Mining: Exploring Hyperlinks, Contents and Usage Data, 2<sup>nd</sup> Edition*. Springer. Verlag Press.
- Liu, B., Zhao, K., Benkler, J. & Xiao, W. (2006). Rule Interestingness Analysis Using OLAP Operations. In Proc. of the 12<sup>th</sup> ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining (KDD '06), pp.297-306.
- Luck, D. & Lancaster, G. (2003). E-CRM: Customer Relationship Marketing in the Hotel Industry. *Managerial Auditing Journal*, 18 (3), pp.213-231.
- Malhotra, N.K. (2010). *Marketing Research; An Applied Orientation 5<sup>th</sup> Edition*. Prentice Hall.
- Mihai, I. (2009). Web-Mining in E-commerce. *Annals of the University of Oradea, Economic Science Series*, pp.959-962.
- Naylor, R.W., Lambertson, C.P. & West, P.M. (2012). Beyond the “Like” button: the impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing*, 76 (6), pp.105-120.
- Ngai, E.W.T. (2005). Customer Relationship Management Research (1992-2002): An Academic Literature Review and Classification. *Marketing Intelligence & Planning*, 23 (6), pp.582-605.
- Paiva, E.L. Roth, A.V. & Fensterseifer, J.E. (2002). Focusing Information in Manufacturing a Knowledge Management Perspective. *Industrial Management & Data Systems*, 102(7), pp.381-389.
- Paulissen, K., Mills, K., Brengman, M., Fjermestad, J. & Romano, N.C. Jr. (2007). Voids in the Current CRM Literature: Academic Literature Review and Classification (2000-2005). In Proc. of the 40<sup>th</sup> Annual Hawaii International Convergence on System Sciences.
- Pfeiffer, M. and Zinnhauer, M. (2010). Can Old Media Enhance New Media? How Traditional Advertising Pays off for an Online Social Network. *Journal of Advertising Research*, pp.42-49.
- Ranjan, J. & Bhatnagar, V. (2011). Role of Knowledge Management and Analytical CRM in business Data Mining based framework. *The Learning Organization*, 18 (2), pp.131-148.
- Romano, N.C. & Fjermestad, J. (2002a). Electronic Commerce Customer Relationship Management: An Assessment of Research. *International Journal of Electronic Commerce*, 2, pp.61-113.
- Romano, N.C. & Fjermestad, J. (2000b). Electronic Commerce Customer Relationship Management: A Research Agenda. *Information Technology and Management*, 4, pp.233-258.
- Shaw, M.J., Subramaniam, C., Tan, G.W. & Welge, M.E. (2001). Knowledge Management and Data Mining for Marketing. *Decision Support Systems*, 31, pp.127-137.
- Song, Q., Shepperd, M., Cartwright, M. & Mair, C. (2006). Software Defect Association Mining and Defect Correction Effort Prediction. *Software Engineering, IEEE Transactions*, 32 (2), pp.69-82.
- Srivastava, J., Cooley, R., Deshpande, M. & Tan, P.N. (2000). Web Usage Mining: Discovery and Applications of Usage Patterns from Web Dat. In Proc. of the SIGKDD Explorations, 1 (2), pp.12-23.
- Swift, R.S. (2001). *Accelerating Customer Relationship Using CRM and Relationship Technologies*. Prentice-Hall, Englewood Cliffs, NJ.
- Tiwana, A. (2001). *The Essential Guide to Knowledge Management: The CRM and E- Business Applications*. Prentice Hall.
- Tufféry, F. (2007). *Data-Mining et Statistiques Décisionnelles : l'Intelligence des Données*. Editions Technip.
- Tufféry, S. (2011). *Data mining and Statistics for Decision Making*. Wiley.
- Turney, P.D. (2002). Thumbs up or thumbs down? Semantic orientation applied to unsupervised classification of reviews. In Proc. of the Association for Computational Linguistics 40th Anniversary Meeting. Association for Computational Linguistics, New Brunswick, NJ.
- Van Wel, L. & Royakkers, L. (2004). Ethical Issues in Web Data-Mining. *Ethics and Information Technology*, 6, pp.129-140.
- Wahlberg, O., Strandberg, C., Sundberg, H. & Sandberg, K.W. (2009). Trends, Topics and Under-researched Areas in CRM Research. *International Journal of Public Information Systems*, 3, pp.191-208.

- Wilcox, S. & Stephen, A.T. (2013). Are close friends the enemy? Online social networks, self-esteem, and self-control. *Journal of Consumer Research*, 40 (1), pp. 90-103.
- Wright, A. (2009). Mining the Web for Feelings, Not Facts. *The New-York Times*, August 23, 2009.
- Xu, M. & Walton, J. (2005). Gaining Customer Knowledge through Analytical CRM. *Industrial Management and Data Systems*, 105 (7), pp.955-971.
- Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y. & Chen, Z. (2009). How Much Can Behavioral Targeting Help Online Advertising. In Proc. of the *Intl. World Wide Web Conference (WWW '09)*, pp. 261-270.
- Yeh, I.C., Lien, C.H., Ting, T.M. & Liu, C.H. (2009). Applications of Web-Mining for Marketing of Online Bookstores. *Expert Systems with Applications*, 36, pp.11249-11256