International Conference on Strategic Innovative Marketing, IC-SIM 2014, September 1-4, 2014, Madrid, Spain

New trends of Intelligent E-Marketing based on Web Mining for e-shops

Jon Kepa Gerrikagoitia[a], Iñigo Castander[a], Fidel Rebón[a]*, Aurkene Alzua-Sorzabal[a]

[a]CICtourGUNE, Donostiako Parke Teknologikoa,Mikekeletegi Paselekua 71 – 3 Solairua. E-20009, Donostia-San Sebastián, Spain

Abstract

E-marketing is the concentration of all efforts in terms of adapting and developing marketing strategies into the virtual spaces: web, social media… In an e-commerce site, e-marketing must help consumers in their purchase. This requires precise knowledge of the customer´s preferences. For this reason, holders of e-shops must find out to whom, to what, how and when to refer to the customer, ergo, to know the "*consumer decision journey*" and strengthen their engagement. This analysis is obtained when the customer is visiting an e-shop because (s)he leaves a digital footprint that can be used to understand his/her needs, desires and demands as well as to improve web presence. These data can be used for data mining to understand the e-marketing and selling processes in a better way. In this paper a survey of 86 e-shops in Spain is presented. In the conclusions, some ideas for good e-marketing practices related to the buying behaviour analysis of customers are shown. Hence, new trends in e-marketing are suggested from a strategic, tactical and operational level in which different data mining techniques ease the purchase and the engagement.

*Corresponding author. Tel.: +34-943-010-885; fax: +34-943-010-846.
E-mail address: fidelrebon@tourgune.org

1. Introduction

In the last decade Internet has experienced a 33% high growth in the number of users worldwide. This has led to a greater acceptance of it by consumers, either as an instrument that facilitates the search for information or as a way to an end for the purchase of the products and services ("The internet a decade later", 2013).

This new technological environment has generated a significant change in the behaviour, habits and trends of the consumer and its acceptance by consumers leads to the increase of the *Electronic Commerce* (EC). Defining e-commerce as the commerce in products or services conducted via computer networks such as the Internet (Turban, Lee, King, Liang, & Turban, 2009).

This commercial activity has forced small businesses to develop an active presence of visibility on the Internet (Chen, Chiang, & Storey, 2012) in order to get notoriety in a global market and add value to their traditional business. With the exception that the virtual and traditional customer behaviour are not identical (Eouzan, 2013).

But visibility of Internet does not only give visibility to a company, also allows to: a) to have the business open to the public permanently; b) reduce costs: material stock, logistic personnel, physical space; c) to study the customer better; d) to remove geographic limitations and e) to create brand and engagement.

In Spain, according to the study of the *National Observatory for Telecommunications and the Information Society* (ONTSI) the B2C e-commerce grew by 13% reaching € 12,383M of turnover. This growth's main cause, in the same way as in the world market, was the increase of Internet purchasers in 2012 which matched 15% (15.2M) (ONTSI, 2013).

A closer data in time (2013) and origin (region level) is available through the *National Statistics Institute of Spain* (INE). According to this statistics institute, in 2013 the e-commerce in Spain has increased to around 11 million people who have made any purchase through the web in the last 12 months. This represents 32% of the total population. Moreover, the region of Spain that has used this type of trade the most is the Basque Country with 42%, followed closely by the Community of Madrid with 40% (INE, 2013).

Putting the focus on the case of Basque Country, *Statistic Institute of Basque Country* (EUSTAT) provides some information about e-commerce by data-bank and annual reports (under the information society topic). In this report, there is an interesting section for the e-commerce. The section reflects that 39% of the users that have been recently connected to the internet have made some kind of web based purchase (EUSTAT, 2014).

As it is visible on the data provided by official statistic institutions, e-commerce is a commercial activity on growth and due to its nature. It benefits the economy of a country greatly and especially to small businesses which are committed to its implementation.

As it happens in real sales, e-commerce based selling must be in concordance with the customer´s necessities prior, during and after the sale. This requires precise knowledge of the customer´s preferences. For this reason, owners of sites must find out to whom, what, how and when to refer to the virtual visitor and develop specific marketing strategies for the Internet (Alzua-Sorzabal, Gerrikagoitia, & Rebón, 2014a).

Defining e-marketing as the concentration of all efforts in the sense of adapting and developing marketing strategies into the web environment. E-marketing involves all stages of work regarding a website, such as: the conception, the projects itself, the adaption of the content, the development, the maintenance, the analytic measuring and the advertisement (Strauss, Frost, & Ansary, 2009). In short, e-marketing makes the usage of the Internet to obtain knowledge about customer preferences through interactive website and agents easier (Teo & Choo, 2001).

To keep the customer's attention on the web presence requires to build up a strong customer relationship and to offer services that attract the customer to visit the website frequently and purchase products and services. Therefore, the sales process requires a deep data analysis to know the "*consumer decision journey*" (Gefen & Straub, 2000).

When a customer is visiting a website (s)he leaves a trace of data, called the digital footprint, that can be used to understand the customer needs, desires and demands as well as to improve its web presence.

By the use of this information applied from Web Usage Mining (Arbelaitz et al., 2013) techniques on the digital footprint left by the user during their navigation; knowledge about the behaviour of the customer's purchase is enriched in order to increase his/her engagement. Then, this knowledge has to be converted into intelligence and, if it is possible, an entertaining presentation of the information required by the customer.

In this context, the real situation of e-marketing will be seen through an on-line survey that has been supplied to retailers of Spain for their completion and four in depth interviews. After analysing the state, enriched technical

solutions through the Web Usage Mining process will be proposed and linked to new discovered e-marketing trends. Finally, a few commercial and open source solutions that can help in the creation of concrete solutions for these new e-marketing trends will be displayed.

2. Methodology

In accordance with the meaningful relevance that e-shops have obtained in the last years, it is crucial to conduct successful e-marketing activities in order to maintain or improve the business in these kinds of companies. At this point, it is necessary to suggest a solution from a strategic vision point of view by developing appropriate e‑marketing strategies. The solution derives from a tactic vision by deciding for certain e-marketing trends and at the end an operational vision which is implemented by selected trends & practices (See Fig. 1).

Even so, it is necessary to know the initial state of e-marketing into a sample of the Spanish enterprises dedicated to e-commerce.

Fig. 1. Business vision and e-marketing

The methodology that has been carried out to know the initial state is based on two different manners of obtaining the real state of e-marketing paved by the e-commerce. On one hand, 86 online surveys have been made to some e-shop owners. On the other hand, out of these 86 companies four have been nominated to have an interview in depth. These interviews were conducted via telephone or face-to-face in the client's own company and the online surveys were collected through a digital form made in Google Forms.

Once the surveys were completed, we proceeded to extract the information from them in order to the draw the conclusions that drive the second stage of the process. This second step consists of: a) a reflection by experts on the detected deficiencies and b) the proposal of practical solutions that facilitate the achievement of strategic e-marketing solutions with vision of business and that incorporate techniques of Data Mining.

3. Results and discussion

3.1. Results of surveys and interviews

On one hand, the analysis of the survey data shows that the majority of the e-shop owners questioned are small enterprises with less than 5 employees in the B2C sector (See Fig. 2a) that commodities, quick sale. They are very comparable in number of products and time in the online business (1-2 years) (See Fig. 2c). They also have a maximum number of online orders that (usually) a range from 101 to 1,000 orders a year and yearly revenue of less than 50,000 € (See Fig. 2b). The main languages supported by the e-shops are Spanish and English.
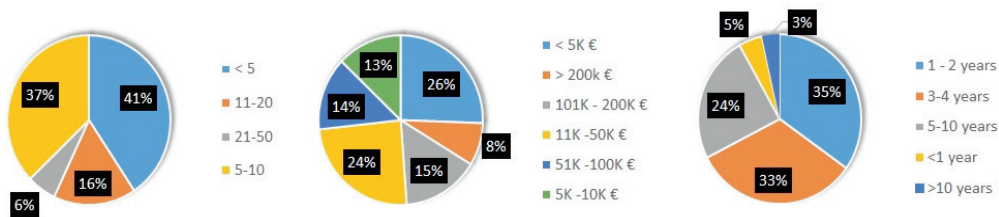

Fig. 2. (a) Number of employees; (b) Annual revenue; (c) Long online business

Although, most of the e-shop owners questioned are currently using a service or tool for customer behaviour analysis in an automatic and online way, in a 40% of the cases more or less they are interested, in a service in a service that is able to analyses the behaviour of groups of customers. 80% of those interested on such a service are aiming to know the customer habits and detect new e-shopping tendencies. Being a 5% of the owners respect to total are conditionally interested. In this case, the requirement is a straightforward system (See Fig. 3a).

The e-shops were also asked about: a) the competitor's observation and the price adjustment and b) the antifraud. In the first case, they said the price is compared to the main competitors' website. In the second case, the proportion of fraudulent cases of their total transaction volume was lower than 0.1% in 53% of the questioned e-shop owners. Out of them, 32% dealt with online payment fraud: by reviewing the incoming transactions or by transferring risk to a third party (See Fig. 3b).
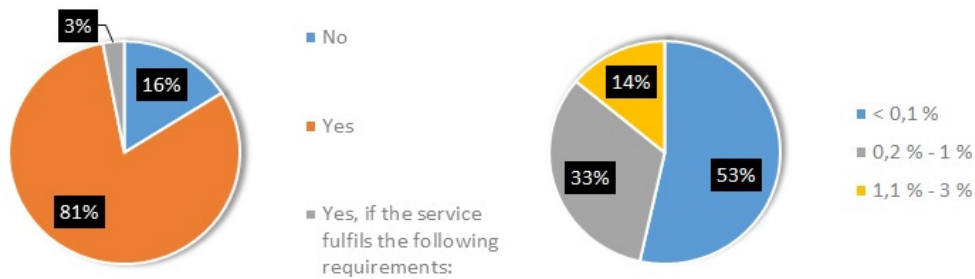

Fig. 3. (a) Interest of behaviour analysis; (b) Proportion of fraudulent cases

On the other hand, analysing the companies that have done the interview, it seems that all of them offer minimally processed, high-demanded products. Furthermore, 50% of the companies surveyed do not know exactly the main reasons why the customers make purchases at their e-Shop. It also is highlights, the fact that only these SMEs employ a person to ensure online sales and make little investments in their online presence (<10K €). In short, these are companies with a small volume of business and their revenues from online sales do not make a big impact on their business.

Remark that, it has been less than 5 years since the e-shops started selling their products via Internet and, therefore, it is a process that must evolve towards maturity. The e-shop considers the on-line service as an added service to its business performance but not a priority.

In addition to this, it can be seen that the e-shops have a certain level of infrastructure because they often have outsourced the hosting of their website and almost all use Google Analytics as a web analytics tool (Clifton, 2012; Hasan, Morris, & Probets, 2009). Out of the e-shops that use Google Analytics, very few do actually take advantage of the tool.

In the same way, according to the surveys, the fraudulent case out of the total volume of transactions does not exceed 1% and they deal with it by leaving it in the hand of third parties: banks and gateways. They do not use electronic tools to observe and compare the prices of its competitors because it is the provider the one that sets them.

To sum up, with the extracted data it is concluded that many e-commerce use Web Analytics tools, specifically Google Analytics, but very few information is obtained from the purpose of the client by his/her digital footprint. This prevents to properly categorize their behaviour and facilitate effective e-marketing campaigns.

3.2. Thinking and discussion based on solutions

The need for developing specific marketing strategies for e-commerce implies that some traditional principles are adapted, or even reinvented. Four activities facilitate the deployment of e-marketing strategies (Stolpmann, 2001): online promotion, online shopping, online service and online collaboration.

In addition to this, we think that it would be helpful it is necessary to choose some of the most relevant e-marketing trends and current trends & practices based on the considerations of experts who are close to the market demand and the practical trends web literature. This means that, a part of scientific base, it has obtained information from other important web origins where it might have significant information about real cases that are closer to reality. Therefore, the description of the trends: brick and mortar, Offering of increasingly complex online features, mobile everywhere... related to the e-marketing is crucial, for then, these can be crossed with trends & practices. These latest trends technological advances must be geared towards a future where retailers can get the most out of their online sales. In this regard, there are certain factors that should be taken into consideration such as the pick-up speed or omnichannel customer experience[2], social-networking testing (Ayada & Elmelegy, 2014), list of wishes, search engine optimization or mega markets[3], cross border e-sales[4], suggestive selling (Mussman, Adornato, Barker, Katz, & West, 2010), web banner advertising (Ozen & Engizek, 2014) and rewards[5].

By doing so, it is obtained the necessary traceability which allows the intelligent e-marketing process be more consistent and stay closer to the e-shops reality.

Table 1. Trends & practices of e-sales according to e-marketing trends.

E-marketing trends	Trends & practices	Activities
Brick and mortar	Technique has not associated trend	Activity has not associated activity
Offering of increasingly complex online features	Pick-up speed	Online shopping & online collaboration
Mobile everywhere	Cross border e-sales	Online promotion
Market consolidation	Web banner advertising	Online promotion & online service
Gamification within e-commerce	Pick-up speed, list of wishes	Online shopping & online collaboration & online promotion & online service
Emotions and loyalty	Social-Networking testing	Online promotion & Online collaboration
Same day delivery	Search engine optimization	Online service
Segment for personalize: the own shop view for each individual customer	List of wishes, suggestive selling, rewards	Online promotion & online service & online shopping

Next, we will introduce the previously described process to facilitate the deployment of e-marketing strategies. The focus will be put on the connection between e-marketing tendencies and the trends and practices of e-sales; which might have a data mining technique implemented in the back-end. Each of them has associated their own activities of e-marketing strategies, being this an important action in order to map both of them (See Table 1).

Moreover, in order to maximise the performance of the e-commerce companies with the use these trends & practices, it is crucial to set up the relationship between e-marketing and the data mining techniques that support the back-end tasks responsible of obtaining a better development of the trends & practices described before. This connection is established in the next matrix (See Table 2). The choice of one technique over another for each of the tendencies has been made according to the way that each trend can demand, and taking into account that the data mining methods can be clustered into two main categories: prediction and knowledge discovery (Han & Kamber, 2006).

Nonetheless, the prediction methods can be categorized as a classification and regression method while knowledge discovery can be included into clustering, mining association rules, and visualization. Despite this, the techniques that these categories offer are too interesting and even more when the data could be analysed by applying

---

2 http://www.practicalecommerce.com/articles/57800-The-Commerce-EvRolution-Part-2-Channel-Trends

3 http://esellermedia.com/2014/01/20/ecommerce-trends-expect-2014/

4 http://www.practicalecommerce.com/articles/4142-Cross-Border-Ecommerce-Booming

5 http://www.practicalecommerce.com/articles/57800-The-Commerce-EvRolution-Part-2-Channel-Trends

data mining techniques and algorithms, in order to identify optimization potential and improve the own marketing and sales processes, the content of web site and e-shop, and the IT-infrastructure (Hassler, 2010). Then, it has been decided to apply the methods that are under these main groups for the diverse trends & practices.

Table 2. Trends and practices according to Data Mining techniques

Trends & practices	E-marketing trends
Pick-up speed	Classification
Social-networking testing	Regression
List of wishes	Clustering
Cross border e-sales	Clustering and/or classification
Suggestive selling	Affinity analysis
Web banner advertising	Clustering
Rewards	Regression
Search engine optimization	Clustering and/or visualization

The mapping in Table 2 provides some suggestions for implementing certain trends & practices as examples. When new trends & practices are addressed, the implementation of data mining techniques needs to be considered.

3.3. Selection Trends and practices based on degree of innovation and in the implementation effort

At this point, an analysis about the diverse tendencies and practices should be made with the objective of knowing what the degree of innovation and the complexity the complexity is supposed to be at Data Mining techniques level, of the development of solution and the time required to completion.

Table 3. Trends & practices based on the degree of innovation and the complexity level

Trends & practices	Degree of innovation	Implementation effort
Pick-up speed	High	Medium
Cross border e-sales	High	High
Web banner advertising	Low	Low
Social-Networking testing	High	High
Search engine optimization	Low	High
List of wishes	Medium	Medium
Suggestive selling	Medium	High
Rewards	Low	Low

According to the values shown in Table 3, note that the assigned values to the degree of innovation and to the implementation effort can be: High, Medium and Low. This kind of qualitative range has been set up, in order to put the focus in a real environment, where the trends and practices can be measured through two variables that provide us which of them might be carried out easier than the others and what trends & practices have a greater innovation than the others in order to add a better value to the e-shops. Therefore, the mapped values that can be seen in the previous table have been established agreeing with the acquired knowledge about the e-marketing and web mining processes.

To do so, the next section will deal with the main Data Mining Suites and open sources tools existing in the market. These kinds of tools facilitate the generation of trends and practices described in this article.

3.4. Commercial and Open Source tools

Currently there is a large volume of vendors and solutions in the … market. In addition to the digital footprint gathering, some vendors are responsible for performing Web Analytics processes. For this purpose, the authors suggest the usage of Piwik (Alzua-Sorzabal, Gerrikagoitia, & Rebón, 2014b).

These gathered data through the digital footprint will be supplied to programs which facilitate the usage of Data Mining techniques. For this reason, in this point is going to describe some data mining commercial products, Data Mining Suites (DMS), focused on the techniques mentioned in the previous step (See Table 4).

For these reasons, and also taking into account one of the most prestigious Data Mining websites' (KDNuggets) Software Poll (14th)[6] and the list of products that appear in the literature (Mikut & Reischl, 2011) or solutions that are more oriented to cloud services systems and *Software as a Service* (SaaS)[7], it can be established the next set of commercial products that might be focused on e-sales.

Table 4. Data Mining techniques according to Data Mining suites

Data Mining techniques	Data Mining suites
Classification	RapidMiner v6 server edition[8], ADAPA[9], SAS Enterprise Miner[10], Skytree server[11]
Regression	RapidMiner v6 server edition, ADAPA, SAS Enterprise Miner, Skytree server
Clustering	RapidMiner v6 server edition, ADAPA, SAS Enterprise Miner, Skytree server
Affinity analysis	RapidMiner v6 server edition, ADAPA, SAS Enterprise Miner, Skytree server
Visualization	RapidMiner v6 server edition, ADAPA, SAS Enterprise Miner, Skytree server

From another perspective, the relative of the open source tools, R and Weka highlight owns. Both of them might implement the data mining techniques that appear in the previous point.

4. Conclusions and future work

This study aimed to test the necessity of the SMEs to become active users of web mining solutions. Overall, from the results it can be extracted that: a) the Pricing analysis is not relevant for them, because in the majority of cases is established by the products providers; b) the fraud has not any impact in the online purchases. It is inferred from the responses, that the market has not obtained the necessary maturity yet and that this issue is often outsourced to third party partners; c) Almost all the questioned companies have deployed the web analytics over their websites but they do not follow an exhaustive data analysis.

Therefore, keeping in mind these key points and observing the significant importance of the digital footprint at the beginning of the web mining processes, it must regard that in a short-term the data origins based on this system will raise and the weight that this step will have on the process will be crucial, in order to dispose a better raw data. Furthermore, as in real sales, the process of selling based on e-commerce must be in concordance with the customer´s necessities prior, during and after the sale. This requires precise knowledge of the customer´s preferences. For this reason e-commerce, must find out to whom, what, how and when to refer to the virtual visitor. In order it will take advantage of applying the data mining techniques in the phase that it corresponds to satisfy the necessity of enrich the data and add value according to better information about the client.

---

6 The 14th annual KDNuggets Software Poll: http://www.kdnuggets.com/2013/06/kdnuggets-annual-software-poll-rapidminer-r-vie-for-first-place.html

7 Cloud Analytics and SaaS Providers: http://www.kdnuggets.com/companies/cloud-analytics-saas.html

8 http://rapidminer.com/news-posts/rapidminer-v6-0/#

9 http://zementis.com/adapa/

10 http://www.sas.com/en_us/software/analytics/enterprise-miner.html

11 http://www.skytree.net/products-services/skytree-server/

Additionally, intelligent e-marketing processes should bear in mind the necessary relationships between e-marketing trends and Data Mining techniques in order to develop specific marketing strategies for the Internet.

By starting to use, the focus is put on marketing campaigns and these are oriented to the necessities of the consumers. If the needs of e-customers would be known the needs of e-customers, the trust knowledge of the relationship between e-commerce actors (consumers and vendors) could be obtained and the tendencies of e-marketing would be closer to the real application of these than the theoretical concept supported by the literature.

The necessity of obtaining customers and overall, the requirement of starting to facilitate the use of these new data economy for the e-commerce entails to obtain receptiveness and opportunity by e-shops.

Finally, providing the tools that add value to the e-marketing campaigns such as the digital footprint monitoring tools, we should introduce in a Web Mining process together with controls where the data mining tools described in this study could be applied. Therefore, in order to facilitate the adoption of effective e-marketing policies, a pilot development with a little volume of e-shops will be proposed, which provides them a solution, composed by several controls or software services oriented to e-commerce. These kinds of controls or services will be enriched with an exhaustive analysis of visitors' typology and when the process is finished it will analyse the acceptation degree among e-shops owners of this new enriched information that is available for them.

Acknowledgements

References

Alzua-Sorzabal, A., Gerrikagoitia, J. K., & Rebón, F. (2014a). Tourism Destination Web Monitor: Beyond Web Analytics. *e-Review of Tourism Research (eRTR)*, *5*(Research Notes), 5. Retrieved from http://ertr.tamu.edu/files/2014/02/enter2014_RN_136.pdf

Alzua-Sorzabal, A., Gerrikagoitia, J. K., & Rebón, F. (2014b). Using MWD: A Business Intelligence System for Tourism Destination Web. *Management*, *2*(1), 62–72.

Arbelaitz, O., Gurrutxaga, I., Lojo, A., Muguerza, J., Pérez, J. M., & Perona, I. (2013). Web usage and content mining to extract knowledge for modelling the users of the Bidasoa Turismo website and to adapt it. *Expert Systems with Applications*, *40*(18), 7478–7491. Elsevier.

Ayada, W. M., & Elmelegy, N. A. (2014). Advergames on Facebook a new approach to improve the Fashion Marketing. *nternational Design Journal*, *2*(2), 139–151. Retrieved from http://www.journal.faa-design.com/pdf/2-2-ayada.pdf

Clifton, B. (2012). *Advanced web metrics with Google Analytics*. John Wiley & Sons.

Eouzan, G. (2013). *Marketing web: Definir, implementar y optimizar nuestra estrategia 2.0*. Ediciones ENI.

EUSTAT. (2014). *Panorama de la Sociedad de la Información Euskadi 2013* (p. 71). Retrieved from http://www.eustat.es/elementos/ele0011200/ti_Panorama_de_la_Sociedad_de_la_Informacion_Euskadi_2013_pdf_962_KB/inf0011206_c.pdf

Gefen, D., & Straub, D. W. (2000). The relative importance of perceived ease of use in IS adoption: a study of e-commerce adoption. *Journal of the Association for Information Systems*, *1*(1), 8.

Han, J., & Kamber, M. (2006). *Data Mining, Southeast Asia Edition: Concepts and Techniques*. Morgan kaufmann.

Hasan, L., Morris, A., & Probets, S. (2009). Using Google Analytics to evaluate the usability of e-commerce sites. *Human centered design* (pp. 697–706). Springer.

Hassler, M. (2010). *Web analytics: Metriken auswerten, Besucherverhalten verstehen, Website optimieren*. Hüthig Jehle Rehm.

INE. (2013). *El comercio electrónico y el uso de las nuevas tecnologías 2012* (p. 6). Retrieved from http://www.ine.es/ss/Satellite?blobcol=urldata&blobheader=application/pdf&blobheadername1=ContentDisposition&blobheadervalue1=attachment;+filename=CifrasINEComopdf.pdf&blobkey=urldata&blobtable=MungoBlobs&blobwhere=972/963/CifrasINEComopdf,1.pdf

Mikut, R., & Reischl, M. (2011). Data mining tools. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, *1*(5), 431–443. Wiley Online Library.

Mussman, D. C., Adornato, R. L., Barker, T. B., Katz, R. A., & West, G. L. (2010). Method and system for providing real time offers to a user based on obsolescence of possessed items. Google Patents.

ONTSI. (2013). *Estudio sobre Comercio Electrónico B2C 2012* (p. 77). Retrieved from http://www.ontsi.red.es/ontsi/sites/default/files/informe_ecomm_2013.pdf

Ozen, H., & Engizek, N. (2014). Shopping online without thinking: being emotional or rational? *Asia Pacific Journal of Marketing and Logistics*, *26*(1), 78–93. Emerald Group Publishing Limited.

Shahabi, C., Zarkesh, A. M., Adibi, J., & Shah, V. (1997). Knowledge discovery from users web-page navigation. *Research Issues in Data Engineering, 1997. Proceedings. Seventh International Workshop on* (pp. 20–29). Birmingham, UK.

Stolpmann, M. (2001). *Online-Marketingmix: Kunden finden, Kunden binden im E-Business*. Galileo Press.

Strauss, J., Frost, R., & Ansary, A. I. (2009). *E-marketing*. Pearson Prentice Hall.

Teo, T. S. H., & Choo, W. Y. (2001). Assessing the impact of using the Internet for competitive intelligence. *Information & management*, *39*(1), 67–83. Elsevier.

The internet a decade later. (2013). . Retrieved July 9, 2014, from http://www.bestedsites.com/the-internet-a-decade-later/

Turban, E., Lee, J. K., King, D., Liang, T. P., & Turban, D. (2009). *Electronic Commerce 2010* (6th ed.). Upper Saddle River, NJ, USA: Prentice Hall Press.

Wang, L., Tao, J., Kunze, M., Castellanos, A. C., Kramer, D., & Karl, W. (2008). Scientific Cloud Computing: Early Definition and Experience. *HPCC* (Vol. 8, pp. 825–830).