A Benchmark for Banks' Strategy in Online Presence – An Innovative Approach Based on Elements of Search Engine Optimization (SEO) and Machine Learning Techniques

Camelia Elena CIOLAC[1]

ABSTRACT

This paper aims to offer a new decision tool to assist banks in evaluating their efficiency of Internet presence and in planning the IT investments towards gaining better Internet popularity. The methodology used in this paper goes beyond the simple website interface analysis and uses web crawling as a source for collecting website performance data and employed web technologies and servers.

The paper complements this technical perspective with a proposed scorecard used to assess the efforts of banks in Internet presence that reflects the banks' commitment to Internet as a distribution channel. An innovative approach based on Machine Learning Techniques, the K-Nearest Neighbor Algorithm, is proposed by the author to estimate the Internet Popularity that a bank is likely to achieve based on its size and efforts in Internet presence.

KEYWORDS: *SEO, Internet Website Popularity, banking industry, Machine Learning, K-Nearest Neighbors*

JEL Classification: C81, O33

Introduction

This paper aims to analyze the state of art in Romanian Internet banking and Romanian banks strategy of using the Internet presence as a distribution channel for their products.

As resulted from the literature review, there is a high potential in Internet as a means of delivering banking products and services and, at the same time, these investments generally have lower costs and higher benefits than the traditional banking channels.

This paper is structured in three parts, following a literature review and the methodology:

1. benchmark for assessing banks' websites performance against underlying web technologies and servers;
2. qualitative analysis regarding Romanian banks' presence on social networks and their openness towards new technologies and services delivered over the Internet
3. proposed quantitative model based on K-Nearest Neighbor algorithm, a

Machine Learning technique, for predicting a bank's Internet Popularity based on its size and effort in Internet presence.

[1] *Camelia Elena CIOLAC,* The Bucharest Academy of Economic Studies, Romania, E-mail ciolac_c@yahoo.co.uk

The contribution of this paper to the scientific progress resides in its novel approach to Internet Popularity assessment, which uses both the banks internal coordinates and the market competition as inputs in a Machine Learning technique. Furthermore, another contribution of the paper resides in the benchmark presented for evaluating banks' content retrieval speed based on technical features that lay beyond a simple interface analysis and can only be discovered through website crawling.

1. Literature Review

A study (Nath et al., 2001) on banks' perception of Internet at the moment when e-commerce was emerging discovered that US banks managers considered this investment mandatory for their bank's survival, but feared that this new banking channel would lead to "a degradation in the customer-bank relationship"(p 35).

Seven years later, an analysis (Murillo et al., 2008) of the factors that influenced online banking adoption among US banks during 2003-2006 identified that smaller banks (also called "community banks") rely more on branch contact with customers where they offer personalized banking services. Regarding the bank's positioning against competitors, the same study uses the Index of Multimarket Contact, the Herfindahl-Hirschman Index and bank-specific measures reaching the following conclusions for US banks: from a financial internal perspective: banks' size (in terms of assets) positively influences the decision to adopt online banking, poor return on asset ratios (less profitable banks) puts pressure for sooner adoption of online banking, while a bad performing loan portfolio negatively affects this decision; from an external perspective: the "adoption by more competitors in relevant markets makes a bank more likely to adopt"(p 17), while a higher market concentration delays the adoption decision; from an organizational perspective: the bank's membership in a banking group increases the adoption speed due to the benefits of knowledge transfer.

In Europe, a relatively similar analysis (Arnaboldi & Clayes, 2010) considers the online banking adoption in banks from several European countries during 1995-2005 and conclude that the bank's business model is a major determinant of adoption: "banks with a comparative advantage in deposit accounts are more likely to automate the easy processing of routine transactions", while "a bank that is more specialized in interest-yielding activities – such as lending –would benefit relatively less of setting up online technologies"(p 9).

Regarding the Internet banking as a distribution channel, it is argued that "communication which must incorporate marketing the internet banking opportunity is crucial for both the attraction and retention of customers" (Weston, 2002). Furthermore, the same study uses the Clarke taxonomy of bank customers' loyalty to the brand and suggests that "contented" customers ("who believe their bank is superior in many ways to alternatives") and "brand champions" ("ambassadors for their current bank") are immediately attracted by the new online distribution channel opened by their bank. Also, through intelligent Internet-based communication, the bank is also able to attract competitors' customers that are either "seducible" ("believe they may be better served by switching to a competitor") or "aggrieved" ("likely to switch to a competitor at the first opportunity"). Previously, another study (Nath et al., 2001) cited Meckbach's discovery that "as a result of positive experiences with online banking, one in six of the bank's new customers are referrals from existing customers". This customer reach opportunity through Internet can be increased even more, by using sponsored search engine results as a tool of gaining more visibility for companies' websites (Jansen & Spink, 2007).

An approach (Dreze & Zufryden, 2003) is to analyze the issue of online visibility, "defined as the extent of presence of a brand, or product, in the consumer's environment"(p 5), based on search engine results, website contents, the website's listing in various online directories and other link referral methods. Still, the approach presented by the study is more suitable for websites belonging to less popular companies/ owners, which is not necessarily the case of Romanian banks.

The value of customer support in Internet banking is essential (Nilsson, 2005), as knowledge on both technology and financial services is required in order to increase customers' use of this banking distribution channel. He also shows the differences between help through seller support in a formal manner and help provided through a social network in a familiar manner, arguing that consumers with little knowledge are more likely to use a social network for learning than contacting the seller support center. From Nilsson's point of view, offering a more pedagogic appearance to the bank's website is likely to have more customers use Internet banking. Furthermore, this idea is shared and extended in other studies(e.g. Chen & Yen, 2004) who argue that playfulness (including branded gadgets to download and use offline from the personal computer), website content's connectedness with the real world(through video& audio media) and the choice for customization can also reflect the quality of Internet presence.

As the website design is an important factor in Internet banking interaction, some research studies(Wenham & Zaphiris, 2003) tested various methods for assessing the usability of these services and concluded that the Heuristic Evaluation is a more comprehensive tool to identify usability problems in the website. Website navigability, or the guiding to finding the desired information in a website, is another important feature that stirs researchers' interest. From the navigability perspective, the ease of search for content is important as "the fewer clicks necessary to find the object required, the greater the navigability and the greater the user satisfaction (Hernandez-Ortega et al., 2007).

A qualitative model for assessing customers' satisfaction in using Internet banking is proposed in (Rahim & Li, 2009), using the following factors: user-friendliness, ease of navigation, customization, website appearance, online customer support, support for transactions, accuracy, up-to-date and sufficient information, security, response time, perceived convenience and information tailored to specific needs. A more bank-specific model was proposed in (Guru et al., 2003) and consists of three functional areas (Informational, Transactional & Customer relationship) each one assessed over three degrees of interaction. In terms of customer relationships over Internet, the model proposed in (Guru et al., 2003) has three levels of interaction, ranging from basic features such as electronic mail and feedback forms, to intermediate interaction through advising tools & what-if calculators and culminating with advanced interaction (e.g. videoconferences, service developments).

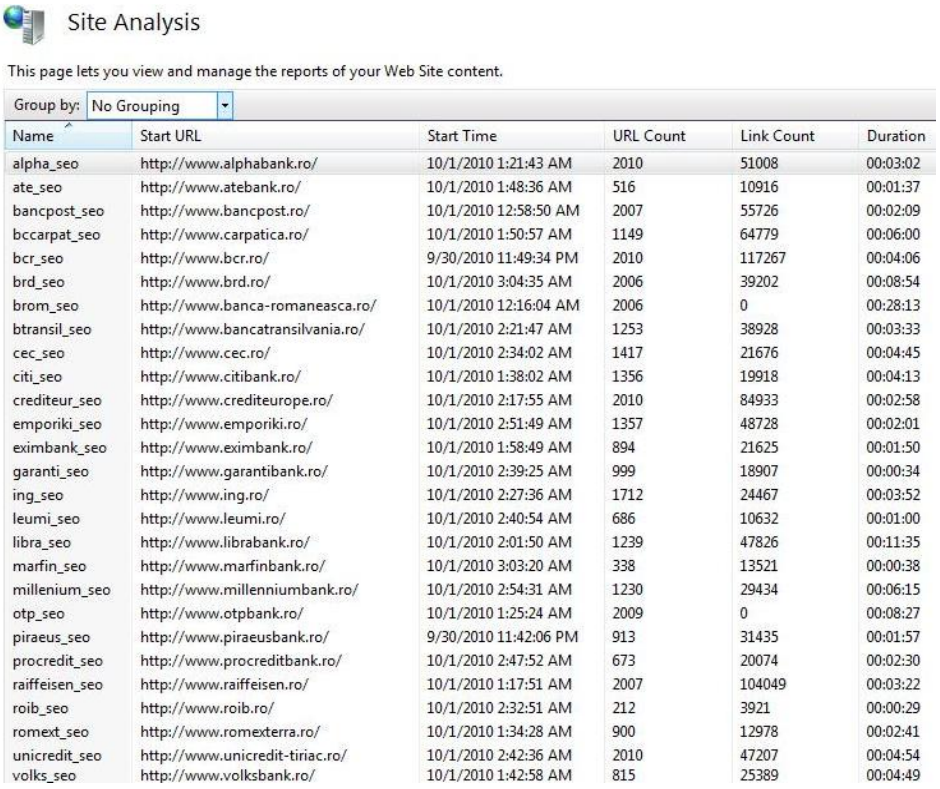
Therefore, from the literature study arise the following keywords that will also be used in our research: Internet presence, bank size (in terms of assets), competitors, reaching the customers, sponsored search engine results, online visibility, website usability, website navigability, response time, support through social networks, pedagogic appearance, connectedness of content.

2. Methodology and Data Set

This paper employs an exploratory approach to the use of Internet as a sales and interaction channel in the Romanian banking industry. The qualitative analysis of Romanian banks' presence on the Internet is doubled by a quantitative study regarding Romanian banks' websites performance in terms of content retrieval speed and Romanian banks' websites traffic ranking.

The data acquisition for the analysis has been collected in a uniform and objective manner using a web crawler, Search Engine Optimization (SEO) Toolkit 1.0 that is part of Microsoft Internet Information Services (IIS) Manager v6.0. The role of this web crawler is to navigate through the structure of a website and collect performance-related data in order to help website administrators improve their online presence and compliance with standards.

For our analysis, a data set of 27 representative Romanian banks has been selected, which is presented in figure 1. For each bank, a standard analysis has been performed in Search Engine Optimization (SEO) Toolkit 1.0 with the following parameters: maximum number of URLs=2000 and maximum download per URL=9999 KB (this being the maximum value allowed by the application).


Site Analysis

This page lets you view and manage the reports of your Web Site content.

Group by: No Grouping

Name	Start URL	Start Time	URL Count	Link Count	Duration
alpha_seo	http://www.alphabank.ro/	10/1/2010 1:21:43 AM	2010	51008	00:03:02
ate_seo	http://www.atebank.ro/	10/1/2010 1:48:36 AM	516	10916	00:01:37
bancpost_seo	http://www.bancpost.ro/	10/1/2010 12:58:50 AM	2007	55726	00:02:09
bccarpat_seo	http://www.carpatica.ro/	10/1/2010 1:50:57 AM	1149	64779	00:06:00
bcr_seo	http://www.bcr.ro/	9/30/2010 11:49:34 PM	2010	117267	00:04:06
brd_seo	http://www.brd.ro/	10/1/2010 3:04:35 AM	2006	39202	00:08:54
brom_seo	http://www.banca-romaneasca.ro/	10/1/2010 12:16:04 AM	2006	0	00:28:13
btransil_seo	http://www.bancatransilvania.ro/	10/1/2010 2:21:47 AM	1253	38928	00:03:33
cec_seo	http://www.cec.ro/	10/1/2010 2:34:02 AM	1417	21676	00:04:45
citi_seo	http://www.citibank.ro/	10/1/2010 1:38:02 AM	1356	19918	00:04:13
crediteur_seo	http://www.crediteurope.ro/	10/1/2010 2:17:55 AM	2010	84933	00:02:58
emporiki_seo	http://www.emporiki.ro/	10/1/2010 2:51:49 AM	1357	48728	00:02:01
eximbank_seo	http://www.eximbank.ro/	10/1/2010 1:58:49 AM	894	21625	00:01:50
garanti_seo	http://www.garantibank.ro/	10/1/2010 2:39:25 AM	999	18907	00:00:34
ing_seo	http://www.ing.ro/	10/1/2010 2:27:36 AM	1712	24467	00:03:52
leumi_seo	http://www.leumi.ro/	10/1/2010 2:40:54 AM	686	10632	00:01:00
libra_seo	http://www.librabank.ro/	10/1/2010 2:01:50 AM	1239	47826	00:11:35
marfin_seo	http://www.marfinbank.ro/	10/1/2010 3:03:20 AM	338	13521	00:00:38
millenium_seo	http://www.millenniumbank.ro/	10/1/2010 2:54:31 AM	1230	29434	00:06:15
otp_seo	http://www.otpbank.ro/	10/1/2010 1:25:24 AM	2009	0	00:08:27
piraeus_seo	http://www.piraeusbank.ro/	9/30/2010 11:42:06 PM	913	31435	00:01:57
procredit_seo	http://www.procreditbank.ro/	10/1/2010 2:47:52 AM	673	20074	00:02:30
raiffeisen_seo	http://www.raiffeisen.ro/	10/1/2010 1:17:51 AM	2007	104049	00:03:22
roib_seo	http://www.roib.ro/	10/1/2010 2:32:51 AM	212	3921	00:00:29
romext_seo	http://www.romexterra.ro/	10/1/2010 1:34:28 AM	900	12978	00:02:41
unicredit_seo	http://www.unicredit-tiriac.ro/	10/1/2010 2:42:36 AM	2010	47207	00:04:54
volks_seo	http://www.volksbank.ro/	10/1/2010 1:42:58 AM	815	25389	00:04:49

Figure 1. "Screen capture from Search Engine Optimization (SEO) Toolkit 1.0 presenting the 27 Romanian banks selected for the analysis"
Source: author

The web crawling procedures have been carried out during nighttime, on the same night (from 30.09.2010 time 11:49:34 PM until 01.10.2010 time 03:04:35 AM), in order to get a better and uniform insight, with no influence from day load activity on the servers. The various categories of results have been extracted and persistently stored in a database for further analysis.

For an accurate interpretation of further results, we must specify the client-side context of the data gathering: a Realtek RTL8139/810x Family Fast Ethernet NIC, Internet Connection type ADSL, Internet Connection Speed 100Mbps provided by the national fix telecommunication provider. The configuration used in this empirical study represents a usual one for a Romanian Internet banking user.

3. A Benchmark for Assessing Banks' Websites Performance against Underlying Servers and Technologies

Following the methodology presented in the previous section of this paper, we collected data from the websites of the 27 banks. The analysis revealed 19 different combinations of web technologies and servers. For each combination, we computed the average speed of download, measured in Kbps, for contents of following types: HTML, Text (PDF and MsWord documents) and Graphics (in JPEG, PNG and GIF formats).

Our analysis continued with a K-means clustering algorithm of these web technologies & servers that revealed their grouping in terms of speed performance. Choosing k=4 and a number of 50 iterations we obtained four clusters with the following features and members (see table 1). This benchmark is useful for decision makers because it provides guidelines for predicting their websites' performance.

Based on the results illustrated in table 1 we make the following observations:

- Upgrading to a newer version of web technologies and server does not necessary result in better performance in all categories of content (HTML, Text, Graphics) compared to the previous version;
- Using a mix of web technologies (e.g. ASP.NET & PHP) instead of one does not make a significant difference in content retrieval speed;
- Knowing the average speed of web content transfer for a web technology/server configuration can be useful in designing the website structure.

For example, considering a PHP v4.3.1/APACHE Server v1.3.29 as a web technology/server configuration for the Internet website of a bank, loading a page with 1MB of graphics contents and 500 KB of HTML contents results in an average time of 30.14+19.40 =49.54 seconds. Considering their clients' perspective, banks can use a caching strategy to increase response time for frequently requested contents and avoid refreshing repetitive contents by using frames. Also, slow pages should be redesigned, accompanied by the introduction of a custom search engine in the bank's website to allow faster retrieval of desired contents without wait times on each intermediary page between the home page and the page of interest.

Table 1. "Results of k-means clustering algorithm (with k=4 and 50 iterations) for web technologies and servers in terms of response speed"

Cluster nr:	**Cluster 1**	**Cluster 2**	**Cluster 3**	**Cluster 4**
Features:	Means	Means	Means	Means
Speed_HTML (Kbps)	25.7637	25.66612	62.54000	1.9900
Speed_Text (Kbps)	120.7554	74.21562	87.39250	460.5000
Speed_Graphic (Kbps)	33.9672	32.86334	32.54304	149.7500
Members (Web Technology / Server):	ASP.NET v2 / Microsoft IIS 6.0	ASP.NET & PHP / APACHE v2.2.3	ASP.NET & PHP / APACHE	PHP v5.2.3 / APACHE v2.2.6
	Java Servlets / IBM HTTP Server	HTML & Tralix / APACHE	JSP v2.1 & Servlet v2.5 / APACHE	
	PHP v4.3.1 / APACHE v1.3.29	Java Servlet si Struts / IBM Websphere	PHP v5.2.9 / APACHE v2.2.11	
	PHP v5.2.6 / APACHE	JSP/ OPENCMS v7.5.2	ASP.NET v2 / Microsoft IIS 7.5	
	PHP v5.3.1 / APACHE	JSP / APACHE v2.0.52		
		PHP v5.1.6 / APACHE		
		PHP v5.2.12 / APACHE		
		HTML / APACHE		

Source: author

A particular attention is drawn by a Romanian bank that uses separate servers for storing PDF documents (on a JETTY v6.1.24 Server) and for the rest of the website (on a OpenCMS v7.5.2). The performance results of this architecture indicate average speed for all types of content: HTML, text and graphics.

Another good practice identified through this analysis is the use of a content management solution. Examples of such solutions identified by crawling the Romanian banks websites include: Avandor Content Management Solution and DotNetNuke.

In terms of operating systems, the data gathered with Search Engine Optimization (SEO) Toolkit 1.0 contained information about the operating systems used by 10 banks. For the rest of the banks this data was not available due to remote server security policies. Out of these, as shown in figure 2, UNIX, Red Hat Linux and Windows Server 2003 were among the most frequently used. Still, it is impressive the large proportion regarding the use of various Linux distributions (Red Hat, Debian, CentOS, Ubuntu): half of the ten banks used this type of operating systems, which is open-source.

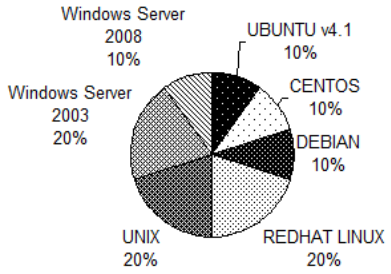


Figure 2. "Structure in the use of operating systems that support banks' websites, data collected from 10 Romanian banks".
Source*:* author

In terms of IT investments for banks, the Apache server license is free for commercial use, Microsoft IIS is part of the operating system and its license cost is included in the operating system license costs, IBM HTTP Server is free for use and the support from the vendor is charged separately.

4. An Analysis of Romanian Banks' Presence on Social Networks

The relations with clients underwent strong transformations after the Internet boom. New communication tools are available and can be exploited successfully to gather feedback from clients, to get a better insight of customers' expectations and needs as well as to promote the bank's financial products.

Even if the SEO-based analysis revealed an openness in using e-mail for communication, as a range of 2 to 1476 mailto protocol usages were identified in each bank website, the real time communication in a public social network or in private is likely to bring more value to the client-bank relationship, especially among younger customers. The messenger protocol, ymsgr, was discovered only on one bank website and its purpose was to facilitate the sending of a video over Yahoo Messenger.

The Romanian banks adhered to the professional network Linkedin in a large proportion. As for the more informal social networks, like Facebook or Netlog, banks interact generally with young customers in order to try to better understand their needs regarding banking products and collect feedback.

On Netlog, only BCR and ING Bank Romania had profiles on the 2nd of October 2010, when this research was carried out, and both of them intensively promoted their students offer and cards offer in their Netlog accounts. Also, various quizzes and polls were opened for visitors. On the other hand, on Facebook, some Romanian banks take advantage of their online presence to stress upon their social responsibility actions, give news about the territorial units extension and new agencies, products promotions, as well as to provide a brief overview of key financial performance indicators of the bank.

Regarding the usage of Twitter, Romanian banks are even more reluctant to possible benefits. The few banks represented on Twitter use it either as a notification channel for the exchange rates & new product offers, or in a true conversational manner to support and advice customers in different banking problems. An example of good practice in using Twitter is ING Bank (http://twitter.com/webcard_ro) that acts as a real-time interaction & counseling tool between customers and bank representatives on various banking issues, also used to test the customers attitude before launching a new financial product/service.

From the investments' point of view, adhering to these social networks is free of charge and the only costs involved are related to the human resource allocated to operate on them. The benefits reflect in increased customers' loyalty to the brand, attracting new clients, obtaining feedback, increased brand reputation, one to one and one to many marketing.

5. The Openness of Romanian Banks Towards New Technologies and Services Delivered Over the Internet

This section of the paper aims to present some best practices of Romanian banks that understood the potential of new Internet services and exploited them in order to increase revenues and also their reputation.

Nowadays, Internet services include a wide range of applications that allow users to interact in a virtual world, allow companies to differentiate themselves from competitors and promote strong localization features. All these technologies can be employed by intelligent banks in order to attract new customers, while increasing customers' loyalty to the brand.

Custom Maps (e.g. Google Maps, NAVTEQ Maps) are a powerful tool for establishing locations on a map that can be viewed in a satellite, road or hybrid view. Banks, through their very nature, rely on their network of units and ATMs, so placing them on a map and giving directions to reach these units can increase customers' satisfaction. Moreover, taking advantage of contextual computing, banks could think of advising customers about the nearest bank unit or ATM. From the analysis carried out, it appears that most of the banks prefer a JavaScript or Flash approach to this issue. Only three banks out of a total of 27 analyzed banks use either Google Maps or NAVTEQ to present the positions of their territorial units on a map on their website. In terms of investment costs, using these two technologies is free of charge. Still, the use of these services implies some costs on the bank side, related to their initial configuration and continuous update.

Google AdWords represents an advertising service provided by Google that enables a website to become a Sponsored Link. This means that the website appears in a special area at the beginning of all results of a Google query for one/more keywords. This approach increases the visibility of the website and usually results in higher traffic and in attracting new clients. In terms of investments, this service is charged on a pay-per-click basis, meaning that the bank is charged every time an Internet user clicks on the corresponding sponsored link. The activation fee for this service is 5$.

Google Insight for Search can and should be used in combination with Google AdWords in order to estimate the frequency of searches per keyword in a certain geographic area. For a couple of banking specific keywords, the analysis carried on the 2nd of October 2010 lead to the results presented in table 2.

We can understand the value of this tool beyond the Internet context: it allows banks to shape their product portfolio based on customers' interests in various country regions. For example, we see in table 2 that the top 3 country regions in Romania that generate searches for "European funds" are Bihor, Constanta and Cluj. A bank using this tool can easily identify this information as being valuable for giving more weight for these products in its offer for these regions.

Table 2. "Competition on various banking keywords in Google AdWords in Romania for the period October 2009 –October 2010 (data retrieved from Google Insight for Search & Google Adwords)"

Keyword	Keyword translation in English	Competition	Local monthly searches	Estimated average cost per click in Romania	Top 10 country regions leaders in searching this keyword
„Credit nevoi personale"	"Credit for personal needs"	High	33100	0.14$	Bucuresti
„Credite imobiliare"	"Real estate credits"	High	8100	0.20$	Bucuresti
„Card de credit"	"Credit card"	Medium	22200	0.28$	Bucuresti
„Credite"	"Loans"	Medium	110000	0.23$	Vaslui, Bacau, Brasov, Galati, Suceava, Bihor, Dolj, Constanta, Prahova
„Imprumut"	"Loan"	Medium	33100	0.05$	Cluj, Timis, Bucuresti
„Fonduri europene"	"European funds"	Medium	40500	0.05$	Bihor, Constanta, Cluj, Timis, Iasi, Bucuresti
„Credit ipotecar"	"Mortgage"	Medium	9900	0.22$	Bucuresti
„Depozite"	"Deposits"	Low	49500	0.29$	Vaslui, Bacau, Bihor, Brasov, Dolj, Prahova, Constanta, Arges, Cluj, Iasi
„Dobanzi depozite"	"Deposit interest rates"	Low	22200	0.26$	Bucuresti

Source: author

Among Romanian banks, the use of Google AdWords service is increasing, examples of implementer banks include: Alpha Bank, BCR, BRD, Banca Transilvania, ING Bank, Libra Bank, Millenium Bank, Piraeus Bank. Using Alexa.com as a tool for websites traffic rankings, we discover that 5/8 of Romanian banks that have a popularity ranking better than 1100 in Romania have implemented Google AdWords, as resulted from our research. The popularity ranking computed by Alexa.com for a special country is based on "a combination of average daily visitors and pageviews from users from that country" (Alexa.com).

Representatives from implementer banks acknowledge the benefits of using Google AdWords. For example, BRD representatives name "the efficient and fast communication, a means to sustain sales and attraction of prospect customers" as being benefits of using the service and argue that it offered more gains than the banners on various websites. At first, BRD used Google Adwords for promoting its private pensions offer and the good results determined the bank to further exploit this online tool for other banking products as well. For Libra Bank, the adoption of this service among marketing practices was gradual and they used the conversion rate of online visitors into bank customers as a metric of investment efficiency. Also, the bank perceived as a benefit of investing in Google AdWords, the possibility of "addressing the target audience in a direct and segmented manner". This paper gives these two examples of good practices to illustrate the efficiency of investments in modern Internet services.

SecondLife is a virtual world in 3 dimensions (3D) where users can interact among them, can develop and configure their own worlds. Similar to a real society, this environment allows trades based on virtual money called L$. BCR is the first Romanian bank to open a

virtual unit in Second Life, having as a primary objective training the customers in financial topics. Also the bank organizes various quizzes that imply knowledge about the real bank's products in order to gain L$. Lately, the quiz regards the real estate credits and has as prizes 10000L$ that allow the winner to buy a house in the virtual world. From the investments perspective, the real bank needed to invest a certain amount of L$ to buy the spaces and equipment in the virtual world, considering an exchange rate of approximately 260L$ for 1$ in LindeX. We conclude that BCR is an example of good practice in exploiting the virtual world for enriching the customers' experience. Since the launch of the virtual unit in SecondLife in January 2010, almost 30000 visitors entered it until the 1st of October 2010.

6. A Proposed Model of Internet Popularity for Banks Based on K-Nearest Neighbors Algorithm (A Machine Learning Algorithm)

6.1 Proposed Model Description

We propose a model to evaluate and predict the popularity of a bank based on its features and strategy in Internet presence.

For this purpose, we shall take the bank's net assets' weight in the national banking system (according to the Romanian National Bank's annual last report of 2009) as a measure of bank's size, regardless its shareholder structure (i.e. regardless it has a majority public or private, Romanian or foreign shareholder structure).

The bank's strategy in Internet presence is assessed based on a scorecard presented in figure 3. The scorecard proposed by us contains four dimensions: Communication, Functionality, Orientation and Training. As noticed, this scorecard is mainly focused on informational and customer relationships issues, and only in part on the transactional side (through the "online credits" item under the "Functionality" dimension), because the transactional aspects exceed the scope of this paper.

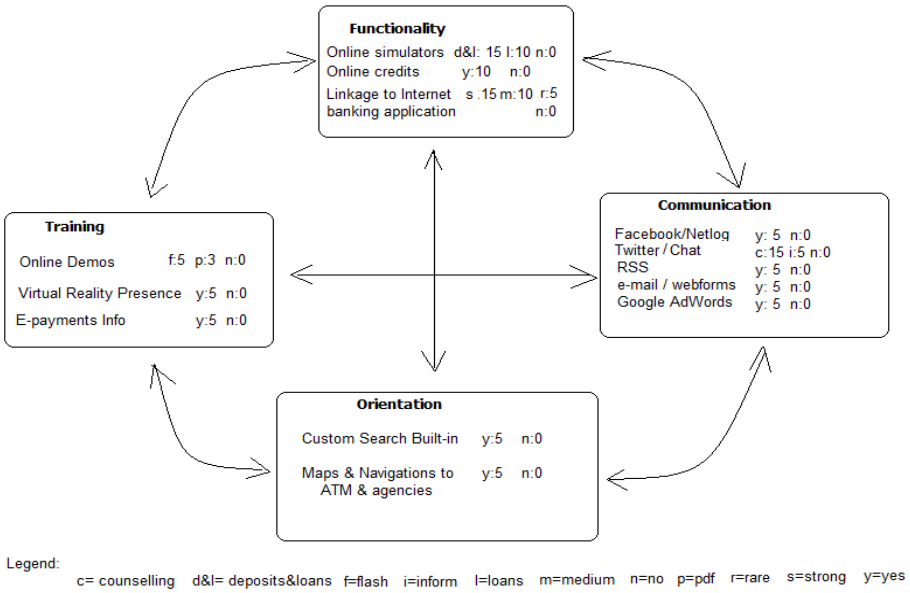


Figure 3. "Proposed scorecard for evaluating banks' strategy in Internet presence"
Source: author

Each of the 27 Romanian banks' efforts in Internet presence has been assessed against the scorecard presented in figure 3. We obtained a wide range of score values between 15 and 80 points. Based on this information we can conclude that the Romanian banking system is heterogeneous in its perception of Internet as a banking distribution channel.

In the next step, for each of the 27 Romanian banks' websites we obtain the Internet popularity as the daily reach counter from Alexa.com. The daily reach counter from Alexa.com reflects the average number of daily visitors from total daily Internet users. We used Internet popularity to classify the banks in one of the four classes:

D: popularity between 0 and 20 visitors per 100000 Internet users
C: popularity between 20 and 40 visitors per 100000 Internet users
B: popularity between 40 and 80 visitors per 100000 Internet users
A: popularity over 80 visitors per 100000 Internet users.

An important aspect should be underlined here: that Alexa.com daily reach counter refers to Internet domains and, from our perspective, this issue is relevant as it incorporates also the number of visitors that use the Internet banking application of the bank.

Next, we apply a Machine Learning technique, the K-Nearest Neighbor in order to simulate the efficiency of different approaches to the Internet presence for banks.

The K-Nearest Neighbor (KNN) algorithm is used for classification and consists of the following steps:

1) A set of m cases of form $(x_i, y_i, c_i)_{i=1,m}$, whose classes $(c_i)_{i=1,m}$ are known a-priori, are displayed on a 2D scatterplot chart in a system of coordinates XOY
2) It is given a new case with an unknown class (u, v, z) where $(u,v) \in$ XOY, z is an unknown class
3) We select a natural number $k \in N$ so that it "achieves the right trade off between the bias and the variance of the model" (Statistica Textbook)
4) A set of k closest cases $(x_i, y_i)_{i=1,k}$ to (u,v) are selected from the 2D plane. The distance used in finding the k closest cases to (u,v) is the Euclidian distance.
5) A voting scheme is applied in order to discover the value of the unknown class z based on the classes of its neighbors $(c_i)_{i=1,k}$

This algorithm is very suitable in estimating benefits correlated with the bank's size and the bank's effort in Internet presence.

Now, from the 27 analyzed Romanian banks, we consider 25 Romanian banks that have a size smaller than 7% of total Romanian banking system's net assets. We chose the threshold of 7% as a selection criterion because the majority of Romanian banks have a weight of net assets in Romanian banking system's total net assets smaller than this value.

We establish the parameters of KNN: m=25 Romanian bank cases of form $(x_i, y_i, c_i)_{i=1,m}$ whose bank sizes $(x_i)_{i=1,m}$, Internet presence effort scores $(y_i)_{i=1,m}$ and classes of Internet popularity $(c_i)_{i=1,m}$ are known a-priori.

From figure 4 we observe that banks with higher scores, i.e. that carried out larger efforts in all the four dimensions of the scorecard (Communication, Functionality, Orientation and Training) generally achieved better Internet popularity than banks of similar size that carried out less efforts in their Internet presence.

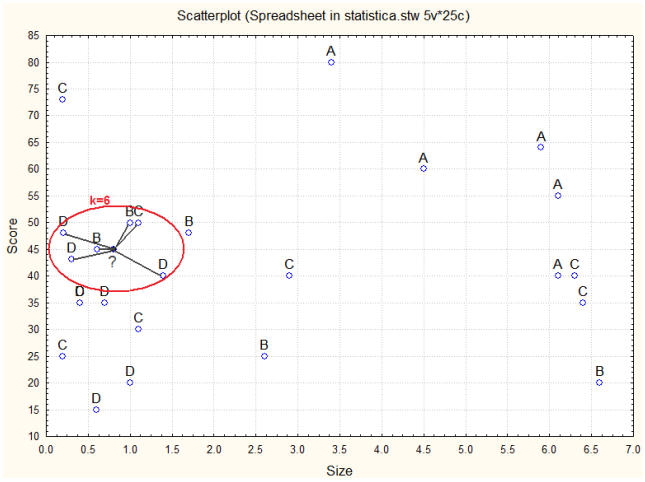


Figure 4. "Simulation of KNN algorithm for k=6 in a 2D scatterplot with m=25 Romanian bank cases known a-priory in terms of size, Internet presence score and class of Internet popularity".
Source: author

Now we consider a new case of a Romanian bank, generically called ODBank, which has size *u*. The bank wants to simulate the potential benefits of various efforts in the four dimensions of the proposed scorecard. Therefore, for various values of v, denoting ODBank's efforts in their Internet presence, the bank computes the K-Nearest Neighbor Algorithm with k between 4 and 6. This way, ODBank is able to assess the optimal benefits it can obtain using the web presence and the underlying effort it must carry out.

6.2 A numerical example for applying the K-Nearest Neighbor (KNN) algorithm in the proposed model of Internet Popularity for Banks

Suppose that ODBank has size u=0.8, i.e the bank's net assets' weight in the national banking system's net assets is 0.8%. This simulation for the case of the fictional bank ODBank is a numerical example and does not reflect a real bank's situation. Performing the computations of KNN algorithm was realized with software Statistica version 7. We applied the KNN algorithm varying the effort in Internet presence(score) of ODBank between 5 and 100 points with a step of 5 points and considering between 3 and 6 nearest neighbors for the voting scheme ($K \in \{3,4,5,6\}$).

Note that the analysis through the voting scheme can lead to the impossibility of establishing a majority if the number of nearest neighbors of two different classes is equal. In table 3, this situation is represented on the vote columns through the symbol n.d. (not decidable).

Considering the example illustrated in figure 4, we see from table 3 that in case ODBank of size 0.8 carries out an effort in Internet presence resulting in a score of 45 points, then the bank can expect a class D of Internet popularity if it compares to 6 nearest neighbors. If ODBank considers that comparing itself to only 3 or 4 other nearest competitors is sufficient, then the bank can expect a class B of Internet popularity following an effort in Internet presence of 45 points.

Table 3. "Results of the simulation of the proposed model for Internet popularity based on KNN algorithm for ODBank"

Score	K=3		K=4		K=5		K=6	
	Vote	confidence	Vote	confidence	Vote	confidence	Vote	confidence
5	D	0.66	n.d	n.d	n.d	n.d	D	0.5
10								
15					D	0.6	D	0.66
20								
25	C	0.66						
30	D	0.66			C	0.6	n.d	n.d
35					D	0.6	D	0.66
40			D	0.75	D	0.8		
45	B	0.66	B	0.5	n.d	n.d	D	0.5
50							B	0.5
55	n.d	n.d			B	0.6		
60			C	0.5	n.d	n.d		
65	C	0.66						
70								
75					C	0.4	n.d	n.d
80	n.d	n.d						
85								
90								
95								
100								

Source: author

The interpretation of the results presented in table 3 shows that, higher efforts in Internet presence can be rewarded with at most a class B Internet popularity of 40 - 80 website visitors per 100000 Internet users daily. Concerning the selection of k value, the analysis shows that considering a higher k results of a wider perspective, including players with similar Internet efforts even if larger in size. Selecting a small k value leads to a tight comparison with 3 other competitors of similar size.

In conclusion carrying out an effort in the four dimensions of the scorecard (Communication, Functionality, Orientation and Training) resulting in a score higher than 50 leads to a C or B class of Internet popularity for the bank. The range of scores 40-50 represent the frontier of two situations for ODBank: carrying out an effort below the score of 50 is very likely to result in a class D of Internet popularity for the bank, with less than 20 website visitors out of 100000 of Internet users. Any score above the level of 50, reflecting higher efforts invested in the Internet presence will lead to a class C or class B Internet Popularity for ODBank. Conversely, the bank can understand in this manner that there is a limit in Internet popularity that cannot be surmounted despite increasing efforts in Internet presence and related IT investments, due to the bank size.

Conclusions

In this paper we analyzed from both a qualitative and a quantitative perspective the Romanian banks' strategy in Internet presence. Using a SEO-based approach we were able to evaluate the performance of various combinations of web technologies, servers and operating systems used. We discovered the little openness of Romanian banks towards new technologies and services delivered over the Internet, even if they are free of charge even for commercial use or provide very high efficiency in terms of benefits and costs.

We proposed a model for banks' websites popularity assessment using a Machine Learning technique, the K-Nearest Neighbor algorithm. In this way, we were able to simulate the Internet Popularity of a bank's website based on the bank's size and efforts assessed against a scorecard proposed by us. The model takes into consideration the results obtained by competitors with similar features with the analyzed bank and involves the k nearest neighbors in a voting scheme to decide the most likely Internet Popularity class of the analyzed bank.

For the future, we aim to continue the analysis by further studying the dependencies between Romanian banks' strategies in Internet presence, their business objectives and the features of target customers.

In conclusion the issues presented in this paper represent a valuable tool for Romanian banks in shaping their strategy for Internet presence. The proposed model can further be applied in other countries.

Acknowledgements

This article is a result of the project POSDRU/6/1.5/S/11 „Doctoral Program and PhD Students in the education research and innovation triangle". This project is co funded by European Social Fund through The Sectorial Operational Programme for Human Resources Development 2007-2013, coordinated by The Bucharest Academy of Economic Studies.

References

Arnaboldi, F. & Clayes, P. (2010). Innovation and performance of European banks adopting Internet, WORKING PAPER SERIES WP 04/10, Centre for Banking Research City University London, Retrieved September 15th 2010, from http://www.cass.city.ac.uk/cbr/Papers/CBR%20WP04-10.pdf

Chen, K. & Yen, D. (2004). Improving the quality of online presence through interactivity, Information & Management 42 (2004), pp. 217–226, doi:10.1016/j.im.2004.01.005.

Dreze, X. & Zufryden, F. (2003). Measurement of online visibility and its impact on Internet traffic, Journal of Interactive Marketing 2004, Retrieved September 20th 2010, from http://xdreze.net/Publications/Visibility_Paper_DZ_1.pdf

Guru, B. K., Shanmugan, B., Alam, N. & Perera, C. (2003). An Evaluation Of Internet Banking Sites In Islamic Countries, Journal of Internet Banking and Commerce vol 8 (no 2), Retrieved September 15th 2010 from http://www.arraydev.com/commerce/jibc/0311-01.htm

Hernandez-Ortega, B., Jimenez-Martinez, J. & Martin-De Hoyos, J. (2007). An Analysis of Web Navigability in Spanish Internet Banking, Journal of Internet Banking and Commerce, vol. 12(no.3), Retrieved September 15th 2010 from http://www.arraydev.com/commerce/jibc/2007-12/Blanca_Final.pdf.

Jansen, B. & Spink, A. (2007). The effect on click-through of combining sponsored and non-sponsored search engine results in a single listing, Workshop on Sponsored Search, Retrieved September 20th 2010, from http://opim.wharton.upenn.edu/ssa3/pdf/submission_96.pdf

Murillo, R. H., Llobet, G. & Fuentes, R. (2008). Strategic online banking adoption, CEMFI Working Paper No. 0813, Retrieved September 20th 2010 from ftp://ftp.cemfi.es/wp/08/0813.pdf

Nath, R., Schrick, P. & Parzinger, M. (2001). Bankers' Perspectives on Internet Banking, e-Service Journal Volume 1, Number 1, Fall 2001, pp. 21-36.

Nilsson, D. (2005). Social Networks And Seller Support In The Internet Bank Use, ANZMAC 2005 Conference: Consumer Behavior, Retrieved September 15th 2010 from http://smib.vuw.ac.nz:8081/WWW/ANZMAC2005/cd-site/pdfs/3-Consumer-Beh/3-Nilsson.pdf.

Rahim, M. & Li, J. Y. (2009). Evaluating an Instrument To Measure Customers' Satisfaction With Internet Banking Applications: A Qualitative Approach, Paper presented at the 8th International Conference on e-Business (iNCEB2009).

Wenham, D. & Zaphiris, P. (2003). User interface evaluation methods for internet banking web sites: a review, evaluation and case stud, in J. Jacko, C. Stephanidis (ed.), *Human-Computer Interaction, Theory and Practice* (pp. 721-725), Lawrence Erlbaum, Mehwah USA, doi=10.1.1.140.4071

Weston, R. (2002). Reaching The Customer - Internet Banking As A Distribution Channel, Academy of Commercial Banking and Finance Proceedings vol 2 Nr 1, Retrieved September 20th 2010, from http://www.alliedacademies.org/public/proceedings/Proceedings10/pacbf-2-1-nash02.pdf#page=48

***(n.d.). Alexa.com. Retrieved October 3rd, 2010, from http://www.alexa.com/

***(n.d.). Google AdWords – BRD. Retrieved October 2nd 2010, from http://www.google.ro/intl/ro/adwords/select/success/brd.html

***(n.d.). Google AdWords – Libra Bank. Retrieved October 2nd 2010, from http://www.google.ro/intl/ro/adwords/select/success/libra.html

***(n.d.). Google AdWords. Retrieved October 2nd 2010, from https://adwords.google.com/o/Targeting/Explorer?__c=1000000000&__u=1000000000&ideaRequestType=KEYWORD_IDEAS#search.none

***(n.d.). Google Insight for Search. Retrieved October 2nd 2010, from http://www.google.com/insights/search/#

***(n.d.). Statistica Textbook. Retrieved October 3rd 2010, from http://www.statsoft.com/textbook/k-nearest-neighbors/