A Study on the Electronic Market for the Successful Launching of a Business

DAN-ANDREI SITAR-TAUT, LIANA-MARIA STANCA, ROBERT BUCHMANN, RAMONA LACUREZEANU
Faculty of Economics and Business Administration, Business Information Systems Department
Babeș-Bolyai University of Cluj-Napoca
Teodor Mihali Street, #58-60 400591, Cluj-Napoca
dan.sitar@econ.ubbcluj.ro
liana.stanca@econ.ubbcluj.ro
robert.buchmann@econ.ubbcluj.ro
ramona.lacurezeanu@econ.ubbcluj.ro

Abstract: - In this article, we will present a study on the Internet market necessary in the case of launching a new electronic business. At present, a large number of e-commerce websites has invaded the virtual space. For this reason, before launching a business, it is very important to make an analysis, especially in the context of Web 2.0 which gives a winning edge over e-businesses that lack social connectivity and high usability features. In this article, we have approached a very important sub-domain of the electronic marketing, namely SEO (Search Engineering Optimization). The launching of an electronic business is very important, but the way the business develops has a vital role, especially the continuous optimization of the dynamic web site, in order to maintain it to the top of the search result list. For our purposes, a statistical study was applied, using the Kolmogorov-Smirnov test.

Key-Words: - Kolmogorov–Smirnov test, search engines, SEO, usability, data mining

1 Introduction
In 1995, the Internet came into the business world in Romania. The companies started to make static web sites, with the purpose of supplying information such as the address of the company and the object of activity. The economic activity was developing in the real world. Soon, these types of web sites lagged behind a new, more dynamic generation. Still, frequent updating behind messages such as under construction” become a disadvantage for the image of the company.

The dynamic web sites were posted on the Internet as a result of the appearance of new technologies that transformed the web into a virtual business environment, dominated by two general notions [9]:
1. content management system – adjustment in real time of the content and structure of the web site;
2. shopping cart or e-shop;
The dynamic web sites represented the starting point in the development of virtual business, electronic education and not lastly, they modified the life of the whole planet. In this context, a theory oriented to the optimal use of these types of web sites has appeared and developed.

2 Problem formulation
The online commerce (e-commerce) represents the selling of products and / or services on the Internet, especially dynamic web sites. The online businesses are not limited only to the selling of various products and / or services on the Internet. An electronic business means much more. It involves the maintenance of the contracts with the suppliers, with the possible business partners, and also with the clients, the promotion of products and / or services offered, etc., all using the electronic resources, dynamic web sites, chats, blogs and e-mails.

The electronic marketing techniques are inspired by the real world and adapted to the virtual environment, in order to generate the highest number of visitors and to make them loyal visitors. Measures have been taken for this and a series of new ideas specific to the virtual environment have appeared, of which the most used is the search engine. The search engine is used as a base element in the promotion of dynamic web sites. The most well known search engines are Yahoo, Google, Ask, Mahalo. As each company wishes to be listed in the top 10 web sites within these search engines, a new research direction hasT developed, namely SEO.

The optimization of web engines and directories is defined according to all the key elements regarded
by the search engines and web directories upon the indexing of the web sites in the database. These key elements, corroborated according to the algorithms of the search engines, offer a certain relevance and importance to the web site according to the information field. According to the relevance and importance factors, the web site will be displayed on a certain position in the list resulting a users’ search.

According to SEO theory, the design and implementation phases of a dynamic web site are based on the principle – the site should allow the user to fulfill his task the best and easiest way possible [10]. The usability is the quality of a system, generally speaking, of a site, in particular, which makes it easy to learn, use, remember, tolerant to errors and agreeable from a subjective point of view [9]. We may conclude that the usability of a dynamic site refers to the relation between the site and the user.

The attributes of a dynamic site for electronic businesses, irrespective of the areas where it is used are: [5]

1. compatibility with all web browsers;
2. obtaining a maximum profit with minimal costs;
3. to be useful;
4. to be easy to learn;
5. to be easy to use;
6. not to include errors;
7. to be able to load any page of the web site any time;
8. to offer safety to the user;
9. to approach interesting subjects;
10. to offer the most recent information, related to the theme of the web site;
11. to include all means of communication adequate to the area of activity.

According to the SEO theory, a basic phase in the preparation of a web site, especially in measuring its usability is the testing phase. In this phase, after the uploading on the web server, it will be verified if [11]:

- the connections within the web site do not lead to non-existent pages and to be placed naturally;
- the site may be accessed from all types of browsers (IE, Firefox) without generating incompatibility errors between the soft versions used to display the information on the computer screen;
- the screen resolution does not modify / distort the structure (design) of the site;
- the loading speed of the site in the browser is the same irrespective of the Internet connection device that the user has (modem, dialup, network);
- entering a web site should not generate errors in javascript or warnings in the status bar of the Internet Explorer browser;
- the content of the web site is analyzed as follows: the purpose of the site, the elements that draw the attention of the users at their first visit, elements that dislike, etc.
- deeper testing will involve traditional software testing techniques:
  a. black box form testing using partitions of input data according to the validation rules;
  b. automated testing using macro recording add-ons for the various browsers (iMacro, Selenium);
  c. gray box testing for the HTML formatting
  d. white box testing if the website was developed with technologies that permit variable watching, code coverage and debugging.

If after this phase the results are satisfactory, the site may be registered in the search engines. Before the registration in the search engines of such site, we think that the preparation of a statistic study of the competition is important. We will make a statistical analysis of the competition web sites within the electronic commerce domain, where we want to launch our business. The study sample is made of 32 web sites that we have chosen from the romanian traffic monitoring site www.trafic.ro, for the electronic commerce domain. We have made an assessment of these web sites based on 10 usability criteria.

### 3 Problem solution

The purpose of the study we have made is the calculation and evaluation made by 13 different evaluators, who have relatively homogeneous background, 8 in each category (advertisement, auctions, commerce and intermediate domains), giving marks for each criterion of the following ten criteria.

- Loading speed
- Communication/feedback instruments
- Transparency
- General aspect
- Publicity
- Deductive interface
- Arrangement of information
- Navigability
- Search engines
- Arrangement of products
- Information quality

For each variable that encodes the marks given for one criterion (discrete, ordinal variable), we have tested the normality condition using the Kolmogorov– Smirnov test [7]. Taking into consideration that this idea has been rejected (p<0.001), non – parameter tests have been applied, and the median, not the average has been used,
considering that this has the capacity to characterize “more cleanly” the variable.

<table>
<thead>
<tr>
<th>Variable</th>
<th>LSpeed</th>
<th>Comm</th>
<th>Transp</th>
<th>GenAsp</th>
<th>Pub</th>
<th>DedIntf</th>
<th>Arrange</th>
<th>Nav</th>
<th>SearchEn</th>
<th>InfQP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Median</td>
<td>2</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
<td>8</td>
<td>9</td>
</tr>
<tr>
<td>95% CI for the median</td>
<td>2 – 2</td>
<td>7 – 8</td>
<td>8 – 8</td>
<td>8 – 8</td>
<td>8 – 9</td>
<td>8 – 9</td>
<td>8 – 9</td>
<td>8 – 9</td>
<td>8 – 9</td>
<td></td>
</tr>
</tbody>
</table>

Kolmogorov-Smirnov test for Normal distribution – Rejected Normality (p<0.001)

<table>
<thead>
<tr>
<th>Sample size</th>
<th>Lowest value</th>
<th>Highest value</th>
</tr>
</thead>
<tbody>
<tr>
<td>416</td>
<td>1</td>
<td>10</td>
</tr>
</tbody>
</table>

Fig. 1 – Kolmogorov-Smirnov test – Normality rejection confirmed

Then, by calculating the Spearman’s rho correlation coefficients [6], we have followed if the official position of the site (given by the number of visits during a determined period) is correlated with the score given by each evaluator, according to the marks provided. We could not find a positive correlation, significant from a statistical point of view, between any variable and the position of the site in the classification.

<table>
<thead>
<tr>
<th>Variable Y</th>
<th>Position</th>
</tr>
</thead>
<tbody>
<tr>
<td>Variable X</td>
<td>LSpeed</td>
</tr>
<tr>
<td>Spearman's coefficient of rank correlation (rho)</td>
<td>0.0676</td>
</tr>
<tr>
<td>Significance level</td>
<td>p=0.1685</td>
</tr>
<tr>
<td>95% Confidence Interval for rho</td>
<td>Lower boundary</td>
</tr>
<tr>
<td>Upper boundary</td>
<td>0.163</td>
</tr>
</tbody>
</table>

Fig. 2 - Spearman’s rho correlation coefficients – No correlations between position and variables

We have also tested if there were differences between the medians of the marks given by each evaluator for a certain criterion (Kruskal–Wallis) [6]. We have found out that although the evaluators had relatively homogenous background, there were various levels of severity in the evaluation.

<table>
<thead>
<tr>
<th>Factor codes</th>
<th>Evaluator</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data</td>
<td>LSpeed</td>
</tr>
<tr>
<td>Test statistic</td>
<td>166.897</td>
</tr>
<tr>
<td>Corrected for ties H</td>
<td>177.554</td>
</tr>
<tr>
<td>Degrees of freedom (DF)</td>
<td>12</td>
</tr>
<tr>
<td>p &lt; 0.0001</td>
<td>p &lt; 0.0001</td>
</tr>
</tbody>
</table>

Fig. 3 – Kruskal-Wallis – No concordance between evaluators’ opinions

We have tested if there were differences between the medians of the marks given for a feature, according to the position of the site. These have been presented (between at least two sites), the only exception being represented by LSpeed. However, we cannot say that there has been a clear correlation between the marks and the position of the web site. We have calculated the median scores of all marks – Mn – (for all criteria) given by each evaluator for each site and we looked at the differences between the median Mn according to the position of the site. Even if they have been identified, as shown in the figure below, there was no clear evolution direction of the marks, the differences can be explained (at least in the first part of the classification) by the
The presence of “discrepant” sites (high position – lower marks, for example those on positions 2, 8, 9, 14). If we exclude these four sites, we do not detect significant differences between the medians of the marks given for the first 20 sites (p>0.05).

The fact that there is no accordance between the marks given to the sites and their official position has also been tested by means of data mining tools, supplied by the application of algorithm C 4.5 (J48 in Weka). Even if this algorithm may fit successfully to such a situation, the results were disappointing again, although we have taken into consideration a tolerance of four positions in the case of classification according to the global position and of two positions, in the case of intra-category classifications (advertisement, auctions, commerce, and intermediate). In both cases, we can find out that the first four attributes occupy the same position, according to the relevance (based on the informational gain). The decision trees resulted, even in pruned variant, present too many ramifications and levels.

The result of the statistical study is that we may launch a business in the electronic commerce domain, because there is no website to excel at all usability criteria. The site that we will launch will be created so that to observe the tested usability criteria, as much as possible.

**Fig. 4 – Median evaluators’ site scores spread and their relationship with site’s position**

<table>
<thead>
<tr>
<th>Class Attribute: Position</th>
<th>Class Attribute: PosByCat</th>
</tr>
</thead>
<tbody>
<tr>
<td>Attribute Ranking</td>
<td>Result</td>
</tr>
<tr>
<td>1.</td>
<td>0.0976</td>
</tr>
<tr>
<td>2.</td>
<td>0.0914</td>
</tr>
<tr>
<td>3.</td>
<td>0.0837</td>
</tr>
<tr>
<td>4.</td>
<td>0.0749</td>
</tr>
<tr>
<td>5.</td>
<td>0.0738</td>
</tr>
<tr>
<td>6.</td>
<td>0.0725</td>
</tr>
<tr>
<td>7.</td>
<td>0.0643</td>
</tr>
<tr>
<td>8.</td>
<td>0.0557</td>
</tr>
<tr>
<td>9.</td>
<td>0.0501</td>
</tr>
<tr>
<td>10.</td>
<td>0.0486</td>
</tr>
</tbody>
</table>

**Fig. 5 – Attributes evaluation and (mis)classification by global and category-based position**
4 Conclusion
The virtual businesses in the Romanian virtual space appear in a stunning rhythm, but not many succeed in keeping up with the virtual market. Taking into consideration this reality, we have thought it is necessary to present a methodology for the research of the electronic market that we consider important both for the start-ups in the electronic economy domain and for the electronic businesses reaching maturity. Any e-business launched on the market needs continuous work for the optimization of the web site, so that it appears and remains in top 10 of as many search engines as possible.

According to the study we have made, we reached the conclusion that the launching of an electronic commerce web site is still a good business in Romania, as the local web sites in this domain have not reached maturity and do not focus on SEO techniques or the wishes of their visitors, and they cannot make the visitors their clients.

In the end, for a successful business, we recommend a good strategy of electronic marketing and a continuous effort for the optimization of the electronic commerce site.

Acknowledgement
This paper was supported by Research Project No. 947, ID_2246/2009 Code, and part of PN II Program financed by the Romanian Ministry of Education, Research and Innovation – The National University Research Council.

References: