ON THE ROAD WITH THE ERASMUS IP WISDOM PROJECT - IMPROVING AN ON-LINE BUSINESS BY APPLYING WEB MINING TECHNIQUES

I. Seruca[1], J. Dias[2]

[1]Universidade Portucalense, REMIT & Univ. do Minho, Centro Algoritmi (PORTUGAL)
[2]Finantech – Sistemas de Informação SA (PORTUGAL)

Abstract

The Web can be regarded as the largest database available and presents a challenging task for efficient design and access. Web mining aims at the discovery and analysis of useful information from the Web through the use of data mining techniques.

In this paper we report our experiences in applying web mining techniques to an e-business platform of a retail company. The company related Web content data and its usage are analyzed with data mining tools. The underlying idea will be to address web marketing purposes, suggesting improvements for both the on-line business and the website. The work here reported was developed in the context of an Erasmus Intensive Programme (IP) project.

Keywords: International cooperation, International project, Web mining, Web marketing, e-business.

1 INTRODUCTION

The rapid growth of the Web in the past two decades has made it the largest publicly accessible data source in the world [1],[2]. Web mining aims to discover potential useful information and patterns from the Web [3],[4]. With the ever-increasing demand for Web-enabled management of knowledge, today's organizations have to address the multiple facets of process, standards, technology, data mining, and warehousing management. This requires ICT approaches to provide an integrated interchange of quality metadata that enable organizations to use the Web as a vehicle for obtaining results that can be both content-rich and practical for decision-making situations.

Web mining can be considered a multidisciplinary topic. In order to have valuable insights, make more informed business decisions and gain competitive advantage, e-business organizations are offered a set of techniques and tools for extracting and analyzing web data ([3],[4],[5],[6],[7]). For transforming the data into useful business information, knowledge of both business and ICT is needed. We argue that Web Mining can provide much support for decision-making in e-business organizations and to improve the quality of service provided by these organizations.

In this paper we use an e-business platform for an on-line business and analyse the web content data and its usage with data mining tools. The underlying idea will be to address web marketing purposes, suggesting improvements for the on-line business, by applying web mining techniques. The data mining techniques applied on the data will provide valuable insights and allow the discovery of interesting patterns; the knowledge acquired may be used by the business to take better decisions. Furthermore the e-business platform will be analysed from the “usage perspective”, through the application of web search models and web analytics tools. This will also allow to acquire knowledge about the way users navigate in the e-business platform and acknowledge ways of improving it.

The work reported here in was developed in the context of an Erasmus Intensive Programme (IP) Project, the IP WISDOM (Web Information System Data Organization Modeling) project, where several European higher education partner institutions were involved [8],[9]. The long-run objective of the IP WISDOM is to build an international curriculum in which the partners can subscribe. Each edition of the IP project was, therefore, a part of a long-term project to develop a joint European Curriculum for Web Mining studies.

The paper is structured as follows: Section 2 provides as overview of the web mining concept and identifies the motivation and challenges involved in the process. Section 3 describes the practical case in which the project was based and sets the objectives to fulfill during the project, identified as a set of business questions to address. In Section 4 we present some findings related with the study performed which address the business questions previously identified. Section 5 includes our

suggestions for the e-business as well as for the website improvement, based on the study performed. Section 6 concludes with considerations on the project achievements.

2 WEB MINING OVERVIEW

Web mining (or Web data mining) is the process of discovering intrinsic relationships (that is, interesting and useful information) from Web data, which are expressed in the form of textual, linkage or usage information. The term “Web mining” was first used by [10] and further adopted by many authors focusing on Web data mining ([12],[13],[14],[15],[16],[17]). A key issue underlying the definitions provided is that Web mining is the application of data mining techniques to discover usage patterns from large Web repositories, being a continually evolving area of technology and business practice. It reveals interesting and unknown knowledge about both users and websites which can be used for analysis. It may be used to understand customer behaviour, evaluate the effectiveness of a particular website and help to quantify the success of a marketing campaign.

Web mining is a demanding and challenging task due to the Web unique characteristics as described below. Furthermore, if used to a full extent, it can be applied in three different branches so as to mine web content, structure and usage.

2.1 The Web Unique Characteristics

There is a general consensus that the Web has many unique characteristics, which make mining useful information and knowledge a both needed and challenging task. These characteristics are summarized in [13] as follows:

1 The amount of data/information on the Web is huge and still growing. The coverage of the information is also very wide and diverse. Information on almost every subject can be found on the Web.

2 Data of all types exist on the Web, e.g., structured tables, semistructured pages, unstructured texts, and multimedia files (images, audios, and videos).

3 Information on the Web is heterogeneous. Due to diverse authorships of Web pages, multiple pages may present the same or similar information using completely different words and/or formats. This makes integration of information from multiple pages a challenging problem.

4 A significant amount of information on the Web is linked. Hyperlinks exist among Web pages within a site and across different sites. Within a site, hyperlinks serve as an information organization mechanism. Across different sites, hyperlinks represent implicit conveyance of authority to the target pages. That is, those pages that are linked (or pointed) to by many other pages are usually high quality pages or authoritative pages simply because many people trust them.

5 The information on the Web is noisy. The noise comes from two main sources. First, a typical Web page contains many pieces of information, e.g., the main content of the page, navigation links, advertisements, copyright notices, privacy policies, etc. For a particular application, only part of the information is useful; the rest is considered noise. To perform fine-grained Web information analysis and data mining, the noise should be removed. Second, due to the fact that the Web does not have quality control of information, a large amount of information on the Web is of low quality, erroneous, or even misleading.

6 The Web is also about businesses and commerce. All commercial websites allow people to perform useful operations at their sites, e.g., to purchase products, to pay bills, and to fill in forms. To support such applications, the website needs to provide many types of automated services, e.g., recommendation services using recommender systems.

7 The Web is dynamic. Information on the Web changes constantly. Keeping up with the change and monitoring the change are important issues for many applications.

8 The Web is a virtual society. It is not just about data, information and services, but also about interactions among people, organizations and automated systems. People can communicate with other people anywhere in the world easily and instantly, and also express their views and opinions in Internet forums, blogs, review sites and social network sites. Such information offers new types of data that enable many new mining tasks, e.g., opinion mining and social network analysis.

2.2 Types of Web Mining

Based on the primary kinds of data used in the mining process, Web mining tasks can be categorized into three types [2],[4]: Web structure mining, Web content mining and Web usage mining.

- **Web structure mining:** Web structure mining discovers useful knowledge from hyperlinks (or links for short), which represent the structure of the Web. For example, from the links, we can discover important Web pages, which is a key technology used in search engines. We can also discover communities of users who share common interests. Traditional data mining does not perform such tasks because there is usually no link structure in a relational table.
- **Web content mining:** Web content mining extracts or mines useful information or knowledge from Web page contents. For example, we can automatically classify and cluster Web pages according to their topics. These tasks are similar to those in traditional data mining. However, we can also discover patterns in Web pages to extract useful data such as descriptions of products, postings of forums, etc., for many purposes. Furthermore, we can mine customer reviews and forum postings to discover consumer opinions. These are not traditional data mining tasks.
- **Web usage mining:** Web usage mining refers to the discovery of user access patterns from Web usage logs, which record every click made by each user. Usage data captures the identity or origin of web users along with their browsing behavior at a website.

3 THE CASE: THE E-BUSINESS COMPANY

Marques Soares is a Portuguese retail company that is mainly focused on the clothing business. It offers a wide variety of products and brands that include ready-to-wear garments for men, women, youth and children as well as other products usually found on department stores such as perfumes, electronic appliances, optics and home décor. Marques Soares reaches their customers through its stores, product catalogue magazine and the online store, which can be found at www.marquessoares.pt.

The company has its roots in Porto, where the first store opened in November 1960. At the time this study was performed, the company had 10 department stores in 7 different cities of Portugal. The company had about 70 000 loyal customers coming from three different sale channels: online sales, physical stores or post mail using the product catalogue magazine. The online store was launched in September 2009 and, in March 2012, the sales on this channel represented only 2% of the global company sales (versus 76% for the stores and 22% for catalogue post mail sales). A major goal for this project was therefore to help the company to identify a strategy to improve the sales and marketing for the e-commerce channel.

3.1 Case Data

For the data analysis several data sources were used. These included the website sales data as an excel file, access to the company's Google Analytics account, a Web server log file and the company profiles in social networks such as Facebook and Twitter as well as videos on the company YouTube account.

Interviews with the representatives of the IT and Sales/Marketing departments of the company were also useful to gather additional information about the company business strategy and goals as well as to help to identify the business questions that should be addressed with this study.

The excel file included the online sales data from the period September 2009 until March 2012. The data sheet was composed of 12761 rows, representative of 5278 sales and 1821 different customers. Each row of the datasheet represented sales transaction data, and included details on customer (number, gender, date of birth, location, profession, admission date, type), product purchased (code, store department, product description, unit price, colour, size, brand, etc.) and transaction (payment method, order number, order date, etc.). Google Analytics data included data about the website visitors and demographics of the users of the website. The web server log provided details such as the IP address of the user/customer, the web browser used, page reference and access date/time.

3.2 Business Questions

From a preliminary analysis of the data provided and the interviews performed with the company representatives the following issues were identified as to be addressed by the study:

1 What is the typical customer profile of the website?
2 What are the major store departments/product categories in terms of on-line sales?
3 What brands or products do customers prefer?
4 Are there any typical combinations of products sold together?
5 What is the geographic location of the visitors of the website?
6 What is the correlation between website visitors by region and website sales?
7 How do the website visitors reach the website?
8 How can the website be improved so that it can be more effective, attract more visitors/customers and increase online sales?
9 What can be done to improve the company use of social media?

4 PROCESSING AND ANALYSIS OF THE CASE DATA

In the scope of this project, data mining was performed using MS Excel and SPSS software and mostly covered attempts to find useful information and discover relationships and patterns on the data analyzed and extracted from web page contents. QlikView was used to present some dashboards. Google Analytics was used to mine the web traffic of the website so as to track user activity patterns from usage logs and user interactions with the website. In the remaining sections, we report the major findings with the study performed and how we addressed the business questions previously identified.

4.1 Typical customer profile, major product categories and brands preferred

The majority of the website customers were women (85,3%) versus 14,7% of male customers. Customers were mainly from the 35-39 age group, as shown in Fig. 1.

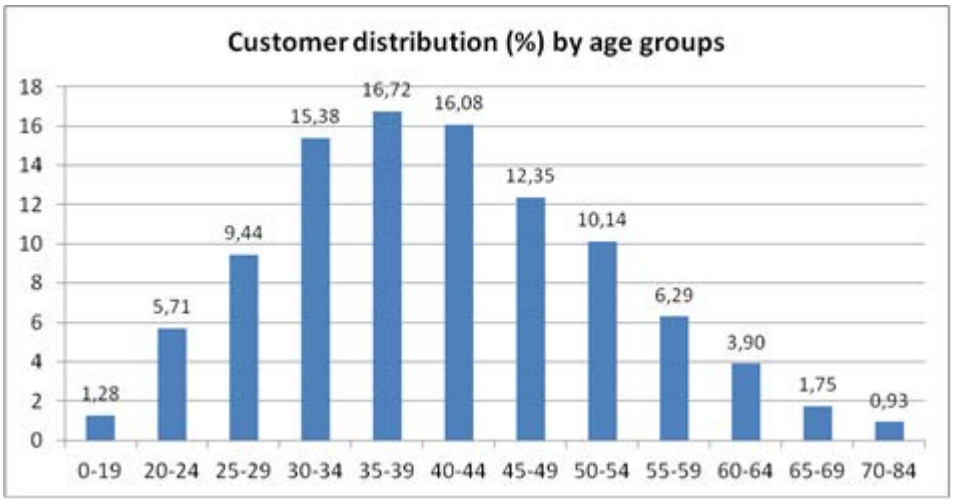


Figure 1. Customer distribution by age groups.

The major product categories in terms of sales revenue were Women, Youngsters and Shoes (Fig. 2), whereas the major product categories in terms of number of sales transactions were Youngsters (23%), Women (19%) and Shoes (14%) (Fig. 3). Both figures show that, in general, clothing sales figures were quite good, while other categories such as accessories, bricolage and optica represented residual sales. It could be wise for the company only to focus on clothing sales.

Store	Mean	Median	Reveneu per product	Nº Sales	% of Total Reveneu	% of Total Nº Sales
Accessories	21,31	18,40	2323	109	,3%	,9%
ANL	77,90	77,90	545	7	,1%	,1%
Beachwear	62,74	65,80	6964	111	,8%	,9%
Bricolage	42,05	20,00	294	7	,0%	,1%
Children	30,01	28,40	31569	1052	3,7%	8,2%
Decoration	19,99	13,75	740	37	,1%	,3%
Electrical apps	159,42	120,00	34436	216	4,0%	1,7%
Home	54,07	45,00	16601	307	1,9%	2,4%
Leather goods	77,88	66,95	12617	162	1,5%	1,3%
Lingerie	26,20	26,65	9798	374	1,1%	2,9%
Men	72,03	55,00	86579	1202	10,1%	9,4%
Optica	143,85	128,50	4316	30	,5%	,2%
Perfumes	56,17	54,90	33817	602	3,9%	4,7%
Shoes	79,58	66,90	145076	1823	16,9%	14,3%
Sports	58,88	57,90	71421	1213	8,3%	9,5%
Watches	91,75	59,00	16883	184	2,0%	1,4%
Women	80,20	68,55	195761	2441	22,8%	19,1%
Youngsters	64,88	59,80	186456	2874	21,8%	22,5%
Total	67,17	56,90	12761	51,795	100,0%	100,0%

Figure 2. Product categories analysis in terms of sales revenue.

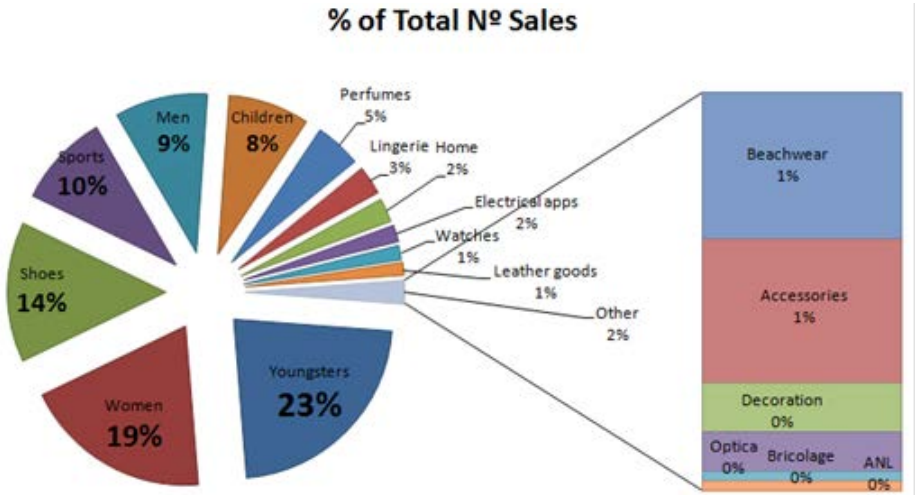


Figure 3. Product categories analysis in terms of sales transactions.

The graph in Fig. 4 shows an overview of the most popular brands sold over the website during the analyzed period of time; “Salsa” is the best selling brand. This information can also be displayed at the year level and month.

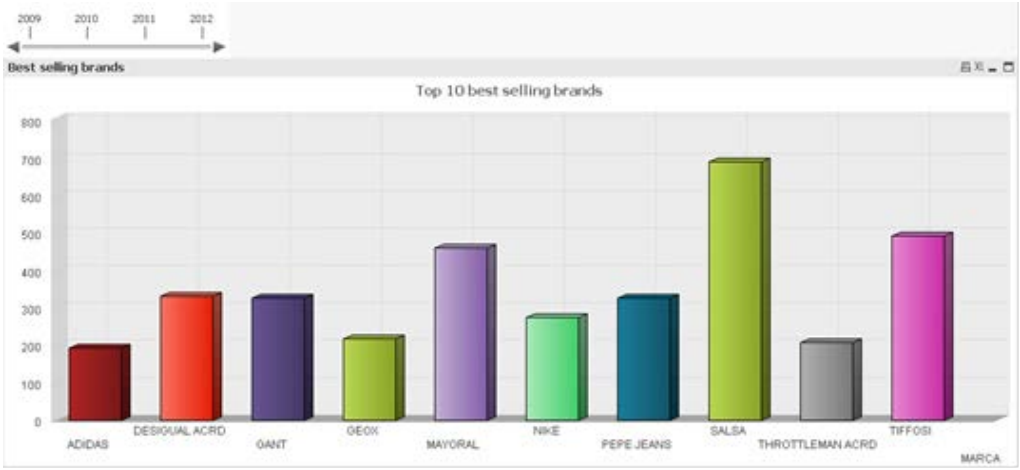


Figure 4.Top 10 best selling brands.

4.2 Combinations of products sold

We performed some market basket analysis experiences so as find products frequently bought together, that could be the basis for a shopping recommendation system. A subset of the results obtained for the product “Bermuda” is shown in Fig. 5.

Selected Item	Bought together	Linked Sales	Avg value for recommendation
Bermuda	Boots	5	59,62
Bermuda	Polo	51	29,33
Bermuda	Trousers	10	33,13
Bermuda	T-Shirt	87	15,65
Bermuda	Sweat-Shirt	16	28,31

Figure 5. Combinations of products sold.

4.3 Geographical location of website visitors

Fig. 6 shows the geographical location of the website visitors, which was ranked by the main cities of Portugal. The cities with more views were Porto, Lisbon and Aveiro.

	Region	Visits	Pages/Visit	Avg. Visit Duration	% New Visits	Bounce Rate
1.	Porto	90,224	11.08	00:05:39	55.15%	14.82%
2.	Lisboa	36,869	10.15	00:06:36	46.49%	19.89%
3.	Aveiro	14,738	11.03	00:05:28	52.87%	17.89%
4.	Évora	9,635	13.10	00:06:56	42.43%	10.05%
5.	Setúbal	8,570	12.22	00:06:38	44.32%	12.95%
6.	Braga	8,360	11.60	00:05:39	57.02%	16.63%
7.	Santarém	5,440	12.66	00:06:41	48.88%	17.52%
8.	Coimbra	4,178	11.84	00:06:26	48.49%	16.44%
9.	Faro	4,163	12.52	00:06:44	48.40%	12.39%
10.	Regiao Autonoma dos Açores	3,846	13.48	00:08:05	38.59%	10.87%

Figure 6. Geographical location of website visitors.

4.4 Correlation of website visits per region and website sales

We also analysed the correlation between the website visits by region and the website sales. These results are shown in Fig. 7. This will enable to have valuable insights of the regions that need to be targeted with improved marketing campaigns.

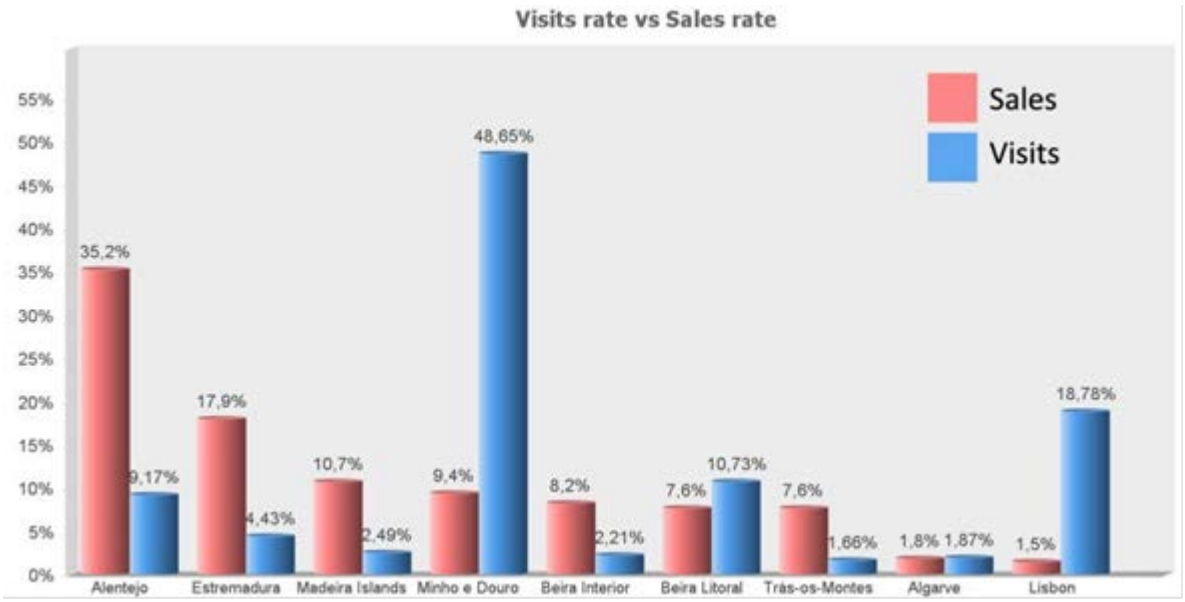


Figure 7. Correlation of website visits per region and website sales.

4.5 Website access sources

As shown in Fig. 8, most of the traffic (87%) came from searches on Google or by direct link access to the website.

1.	google / organic	15,108	66.59%
2.	(direct) / (none)	4,588	20.22%
3.	facebook.com / referral	483	2.13%
4.	bing / organic	313	1.38%
5.	beta.mail.sapo.pt / referral	180	0.79%
6.	mail.sapo.pt / referral	156	0.69%
7.	pesquisa.sapo.pt / referral	123	0.54%
8.	webmail.iol.pt / referral	108	0.48%
9.	marquessoares.pai.pt / referral	74	0.33%
10.	alojadaslojas.com / referral	73	0.32%

Figure 8. Top sources used to get access to the website.

4.6 Website Goal Conversion Rate in Purchases over the Internet

As shown in Fig. 9, the goal conversion rate of the purchases was 3,52%, which means that only that rate of the website visitors actually placed an order. Such a low rate can be caused by several issues. However, a policy definition to offer some kind of benefits to the online customers should be considered.

Figure 9. Goal conversion rate in purchases over the Internet.

5 SUGGESTIONS FOR IMPROVEMENT

From the data and web mining performed, it was clear the company website needed to be improved so that it could be more effective, attract more visitors/customers and, therefore, become a means to increase online sales. The following suggestions for improvement were identified, so that the company could more easily address these goals:

5.1 Search Engine Optimization (SEO) and Website advertisement

Quite often people do not know the exact URL of a website and use a search engine for a prior search. However, from the results obtained with Google Analytics shown in Fig. 10, it is clear that the most used keywords are all related to “Marques Soares”; that means that these users already know the store and probably are customers.

	Keyword	Visits ↓	Pages/Visit	Avg. Visit Duration	% New Visits	Bounce Rate
1.	marques soares	9,588	16.32	00:06:05	43.09%	10.40%
2.	(not provided)	1,965	16.41	00:06:18	51.35%	11.81%
3.	armazens marques soares	257	17.73	00:07:25	47.86%	10.51%
4.	www.marquessoares.pt	195	16.47	00:07:32	37.44%	11.28%
5.	marquessoares	167	15.63	00:08:06	37.13%	16.17%
6.	marques soares catalogo	144	18.73	00:06:10	52.08%	6.94%
7.	marques soares porto	129	15.00	00:04:22	57.36%	17.83%
8.	marques e soares	122	20.80	00:06:40	54.92%	9.02%
9.	http://www.marquessoares.pt/	107	13.79	00:07:22	43.93%	14.02%
10.	www marques soares	83	16.45	00:06:54	50.60%	9.64%

Figure 10. Keywords used in Google for searching the company website.

The company needs to attract new visitors that have never heard about the store. The goal should be to have a higher page rank when users just look for clothing stores. The following hints should be considered:

- *Use better keywords*

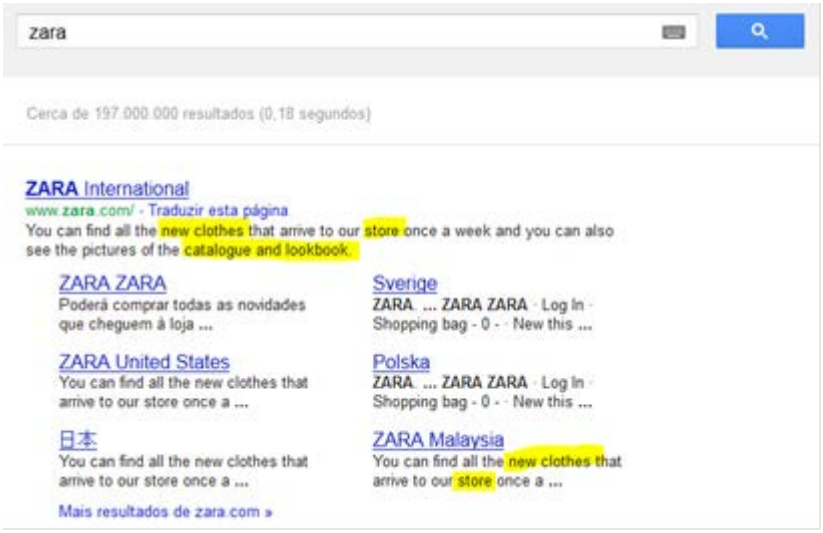
For simple keywords such as “loja de roupas” (clothing store), “sapatos” (shoes) and “calçado” (shoes), Marques Soares’ website does not show up on the first page in Google. Other keywords should be considered such as “pronto-a-vestir” (ready to wear) and “grandes armazéns” (department store).

- *Add metadata*

The use of metadata is a way to make a website appealing to new customers. Metadata contains information about a web page for both search engines and visitors. It allows to set up the description provided by any search engine when a website is shown in the search results. At the time of the study performed, Marques Soares did not have any metadata.

The store “Zara” was one of the closest competitors to Marques Soares in the local market. Hence, a comparison about the results of both searches in Google was made. As shown in Fig. 11, the description of Zara’s webpage was more appealing and clearer for visitors when compared with a similar description about Marques Soares (Fig. 12).

Zara’s metadata: *<meta name="description" content="Poderá comprar todas as novidades que cheguem à loja cada semana e também encontrará as fotografias do catálogo, do lookbook e da colecção." />*



New clothes, catalogue and lookbook, store.
Figure 11. Metadata and results obtained with a search in Google for the Zara store.

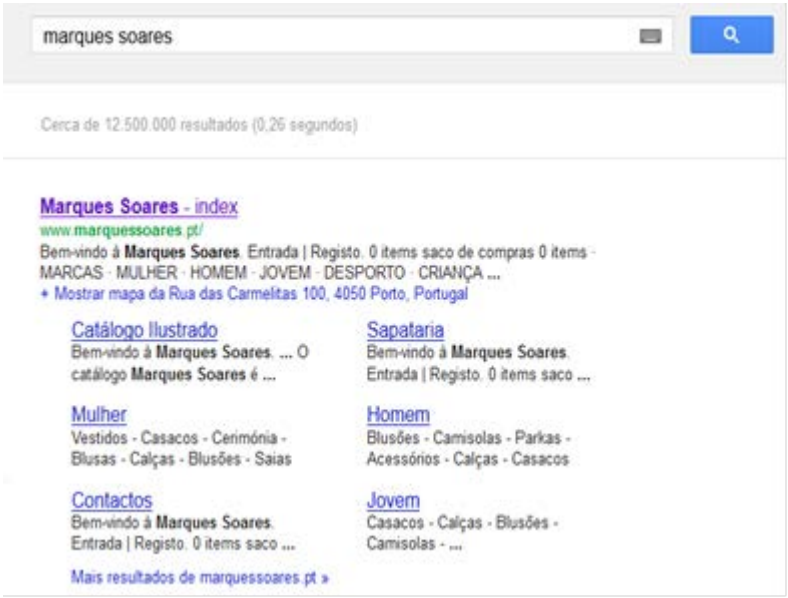


Figure 12. Results obtained with a search in Google for the Marques Soares store (accessed in March 2012).

✦ *Website advertisement*

As displayed in Fig. 8, when considering the access to the website, most of the hits came from "Google", when people were looking for "marques soares". A possibility to further increase the traffic and the number of potential customers would be to optimize the website for other search engines. Furthermore, it is worth noticing that there are no visitors coming from other websites. Therefore, the company should consider creating some ads to be placed on other websites. These online advertisements of the company should target popular sites that are visited by potential customers. Our advice was to first focus to get the website up in the search results, so that the number of visitors and customers may more easily increase. Subsequently, advertisement should be used as well.

5.2 Product Recommendation system

The site didn't have any kind of product recommendation system nor a last/recently viewed items section. Such a system should be useful, not only to offer a personalized marketing and website experience to the customer but also to improve the online sales. Our market basket analysis

experiences showed some interesting results and uncovered the need for the implementation of such a system.

5.3 Promote online sales and reward loyal and good customers

Offer special conditions for online sales, such as a discount rate specially set for online sales and/or offer free shipping for orders above a given amount. Offer the so considered “good customers” vouchers, gift cards or access to private sales; reward a customer when he brings on a new customer.

5.4 Improve the interaction with customers through social media

The level of interaction with customers through the company facebook page was rather low. Thus, interaction with customers through the posting on the facebook page of other contents (e.g. sports events) that may attract the attention of some user groups may be an option. The launching of contests in the facebook page which award winners promotion coupons for use in the store is another possibility. Facebook open social graph meta tags (https://developers.facebook.com) should also be added, so that when people share a product on their facebook wall, the proper photo and description are shown. On the other hand, the Twitter account was merely used as a copy of the facebook content. Twitter may be easily used to spread information such as deals, new stores, new products and ensure a direct communication with the customer. Twitter should also be used to target users following brands that the store sells or that are at the same level of interest such as Levis, Adidas, Burberry, etc.

When the number of followers of both the company pages in facebook and Twitter will reach a higher volume, a monitoring social media tool may be used, to allow the company to effectively analyse what the customers think about the company, its products and the service provided, and further enrich its customer information database with other interesting data, such as the events and places that the customers are interested in, the people they are following, their network of friends, etc. This kind of information should be considered in the launching of future marketing campaigns.

5.5 Make the website accessible for mobile devices

A mobile version of the site for tablets and smartphones should be useful, specially to attract younger customer age groups and for product marketing purposes, as through mobile devices it is easier to share clothing products information.

5.6 Website load time optimization

Loading time is a major contributing factor to page abandonment; slow page response time also results in an increase in page abandonment. To decrease the load time of the website, the following actions were advised to be considered:

- Reduce the number of JavaScript and css files used so as to reduce the number of HTTP petitions. At the time of the study, 5 css and 11 JavaScript files were used.
- Use a cache system (e.g. memcached). The most visited pages such as the index should be in memory, ready to be served.
- Use a Content Delivery Network (CDN) to serve static files; css, images and js files can be served using a CDN, such as Amazon S3. The jquery version cached by Google is another possibility.
- Put the JavaScript files at the bottom of the page. The html content will load faster.
- Minimize image size: the images had a lot of white space resulting in a waste of bandwidth.

6 CONCLUSIONS AND FURTHER WORK

Fundamental to the optimization process proposed in this paper was measurement, gathering data and information that could be transformed into tangible analysis and recommendations for improvement, by using Web mining tools and techniques. We mainly focused on a quantitative analysis of online visitors and customers behaviour, as from the data provided we could not have a more qualitative view of online behaviour so as to report on the overall user experience and report

direct feedback given by visitors and customers, even though we could infer some customer experiences at that level (e.g. a high bounce rate means that the website is not fully optimized).

Overall we felt that we succeeded in our effort of defining a set of recommendations for the e-business as well as for website improvement based on the web mining performed. This opinion was corroborated by the project partners, the other project teams and by stakeholders - the company representatives - who were invited to participate in the project final session in order to evaluate the project outcomes. In spite of this, due to time constraints – the overall project covered 10 working days including teaching sessions – , we didn't manage to further enrich the analysis of data with off-line data from company sales and customers, nor to explore the web mining possibilities to a full extent, which would be valuable directions for further work to consider. Finally, building on the experiences and knowledge gained with the project, a curriculum of a Web Mining subject was set within a postgraduate course in Business Intelligence offered at Portucalense University.

REFERENCES

[1] A. Monelli, S. B. Sriramoju, "An Overview of the Challenges and Applications towards Web Mining", Proceedings of the Second International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)", pp.127-131 Aug, 2018.

[2] K. Griazev, S. Ramanauskaité, "Web Mining Taxonomy", Proceedings of the 2018 Open Conference of Electrical, Electronic and Information Sciences (eStream), 1-4 Apr, 2018.

[3] M. U. Hassan, K. Shaukat, D. Niu, S. Mahreen, Y. Ma, X. Zhao, M.A. Shabir, "Web-Logs Prediction with Web Mining", Proceedings of the 2nd IEEE Advanced Information Management, Communicates, Electronic and Automation Control Conference (IMCEC), pp.1295-1299 May, 2018.

[4] M. P. Bharti, T. J. Raval, "Improving Web Page Access Prediction using Web Usage Mining and Web Content Mining", Proceedings of the 3rd International conference on Electronics, Communication and Aerospace Technology (ICECA), pp.1268-1273 Jun, 2019.

[5] E. Turban, R. Sharda, D. Delen, *Business Intelligence and Analytics: Systems for Decision Support*. 10th Edition, Pearson, 2018.

[6] S. Van Belleghem, *The Conversation Company: Boost Your Business Through Culture, People & Social Media*. London: Kogan Page, 2012.

[7] J. Leskovec, A. Rajaraman, J. D. Ullman, *Mining of Massive Datasets*. Cambridge: Cambridge University Press, 2014.

[8] Wisdom, 2011, Web Information System Data Organization Modeling, project documentation, http://wisdom2011.viko.lt, 2011.

[9] Wisdom, 2012, Web Information System Data Organization Modeling, project documentation, http://wisdom2012.upt.pt, 2012.

[10] O. Etzioni, "The World Wide Web: Quagmire or Gold Mine?", *Communications of the ACM*, 39(11), pp. 65-68, 1996.

[11] M. Zdravko, D. Larose, *Data Mining the Web - Uncovering Patterns in Web Content, Structure and Usage*. John Wiley and Sons, 2007.

[12] P. Lingras, R. Akerkar, *Building an Intelligent Web: Theory and Practice*. Suldbury, MA: Jones and Bartlett Publishers, 2008.

[13] B. Liu, *Web Data Mining – Exploring Hyperlinks, Contents and Usage Data*, Springer-Verlag, 2011.

[14] D. Rathod, "A Review On Web Mining", *International Journal of Engineering Research and Technology (IJERT)*, 1(2), pp. 21-25, 2012.

[15] G. R. Bharamagoudar, G. T. Shashikumar, R. Prasad, "Literature Survey on Web Mining", *IOSR Journal of Computer Engineering*, 5(4), pp. 31-36, 2012.

[16] K. C. Srikantaiah, V. R. Venugopal, *Web Mining Algorithms*. LAP Lambert Academic Publishing, 2014.

[17] R. Gupta, “Journey from Data Mining to Web Mining to Big Data”, *International Journal of Computer Trends and Technology (IJCTT)*, 10 (1), pp. 18-20, 2014.