

The 8th International Conference on Ambient Systems, Networks and Technologies  
(ANT 2017)

## Automating the Extraction of Static Content and Dynamic Behaviour from e-Commerce Websites

João Pedro Dias<sup>a,\*</sup>, Hugo Sereno Ferreira<sup>b</sup>

<sup>a</sup>Artificial Intelligence and Computer Science Laboratory, Faculty of Engineering of University of Porto, 4200-465 Porto, Portugal

<sup>b</sup>Department of Informatics Engineering, Faculty of Engineering of University of Porto, 4200-465 Porto, Portugal

---

### Abstract

E-commerce website owners rely heavily on analysing and summarising the behaviour of costumers, making efforts to influence user actions and optimize success metrics. Machine learning and data mining techniques have been applied in this field, greatly influencing the Internet marketing activities. When faced with a new e-commerce website, the data scientist starts a process of collecting real-time and historical data about it, analysing and transforming this data in order to get a grasp into the website and its users. Data scientists commonly resort to tracking domain-specific events, requiring code modification of the web pages. This paper proposes an alternative approach to retrieve information from a given e-commerce website, collecting data from the site's structure, retrieving semantic information in predefined locations and analysing user's access logs, thus enabling the development of accurate models for predicting users' future behaviour. This is accomplished by the application of a web mining process, comprehending the site's structure, content and usage in a pipeline, resulting in a web graph of the website, complemented with a categorization of each page and the website's archetypical user profiles.

1877-0509 © 2017 The Authors. Published by Elsevier B.V.  
Peer-review under responsibility of the Conference Program Chairs.

**Keywords:** Web Mining; E-commerce; Web Analytics

---

### 1. Introduction

E-commerce, one of the most disruptive innovations in trading that changed the way people buy and sell goods, has a market cap expected to reach  $\approx$  \$2 trillion in 2016, accounting for  $\approx$  9% of total retail sales worldwide<sup>1</sup>.

When a user comes to a given e-commerce website, several marketing and advertising mechanisms are activated that attempt to influence the user actions in order to improve sales and profits. These mechanisms rely heavily on analysing the behaviour of costumers, making a significant impact in the Internet marketing activities.

There are several companies today who develop and provide tools and mechanisms to e-commerce business owners in order to make them capable of improving their business. These tools consist mostly of tracking and learning costumers habits, often being complemented with automatic recommendation systems<sup>2,3</sup>.

---

\* Corresponding author. Tel.: +351 912831605.  
E-mail address: [jpmdias@fe.up.pt](mailto:jpmdias@fe.up.pt)

When faced with a new e-commerce website, the machine learning practitioner typically spends a great amount of time collecting and analysing the static and dynamic data of the website. This process is essential to extract useful information about the website's structure, content, and its users' behaviour. Only after is he/she able to build relevant models and algorithms to enhance on-line marketing activities<sup>4</sup>. This is a great challenge due to the heterogeneous nature of the Web data, the semi-structured or unstructured way that it is presented, and the overwhelming amounts it is being produced<sup>5</sup>.

The goal of this paper is to present a process capable of, given an e-commerce website, extracting useful information from its structure, content and its typical users' behaviour — using state of the art techniques and methods — and returning a consistent model representing the website content, relations between pages and its archetypical users.

In Section II we present an overview of the literature in e-commerce and web mining. Section III gives more details about the problem we are approaching and provides further insight into the proposed solution. Some implementation details are given over Section IV. We conclude with some evaluation in Section V and present some final remarks in Section VI.

## 2. Literature Review

### 2.1. E-commerce

In the past decade, e-commerce became essential for small and large businesses worldwide, not only as a tool for selling goods directly to consumers and other businesses, but also as a way of engaging them<sup>6</sup>.

One cannot talk about e-commerce without mentioning Amazon<sup>1</sup> and eBay<sup>2</sup>, which were among the first big Internet companies to bring widespread e-commerce to the users. But, as of today, these kind of transactions are pervasive on the Internet, ranging from within social networks like Facebook<sup>3</sup>, to *in-app* purchases<sup>7</sup>. It is now increasingly common to find marketing and advertising campaigns on the Internet, with e-commerce companies resorting to Web personalization techniques as a way of doing target marketing over its visitors with the goal of improving the likelihood of a visitor generating profit<sup>8,9</sup>.

In order to measure the engagement and success of an e-commerce website, service owners resort to *e-metrics*<sup>10–12</sup>. These metrics attempt to provide a better and more direct overview of the online business performance, namely:

- **Bounce rate** is the rate of users that visit a single page and jump off the website;
- **Conversion rate** is the number of buyers (or paying customers) over the total number of website users;
- **Click-through rate** stands for the ratio between the users who click on a specific link to the total number of page visitors;
- **Costumer retention** shows the ability of a website in retaining customers on the long term, and;
- **Shopping cart abandonment rate** gives the number of times in which an item was added to the shopping cart but the order was not completed.

One of the core tools to improve the *e-metrics* are *recommendation systems*. These systems rely on a three-step process, starting with (i) data retrieval, normally resorting to Web mining techniques, where we get to know the user preferences by analysing static content and user behaviour data, followed by (ii) computing and validating the recommendation using proper techniques, and finalizing by (iii) presenting the recommendation results to the customer<sup>13</sup>. These systems commonly can make use of three different approaches: collaborative filtering, content-based filtering, and hybrid.

Despite the heterogeneity of the Web one can infer a common structure of e-commerce websites due to their common purpose, *viz.* (i) displaying products to a costumer, (ii) giving them the possibility of navigating between different products, and (iii) allowing purchases. This common structure and design — *de-facto* guidelines followed by websites owners — help the user to easily adapt to any e-commerce website, enhancing its overall experience<sup>14</sup>.

### 2.2. Web Mining

Data mining has been used by companies in order to focus on most important information present in their data. Data mining communities identify three different types of mining: data mining, Web mining, and text mining<sup>15</sup>. Web mining conciliates data mining and text mining techniques applying this concept to the WWW<sup>16</sup>.

<sup>1</sup> <http://amazon.com>

<sup>2</sup> <http://ebay.com>

<sup>3</sup> <http://facebook.com>

Download English Version:

<https://daneshyari.com/en/article/4961176>

Download Persian Version:

<https://daneshyari.com/article/4961176>

[Daneshyari.com](https://daneshyari.com)