

Automatic detection of e-commerce availability from web data

Blazquez, Desamparados^a; Domenech, Josep^a; Gil, José A.^b and Pont, Ana^b

^aDepartment of Economics and Social Sciences, Universitat Politècnica de València, Spain

^bDepartment of Computer Engineering, Universitat Politècnica de València, Spain.

Abstract

In the transition to the digital economy, the implementation of e-commerce strategies contributes to foster economic growth and obtain competitive advantages. Indeed, national and supranational statistics offices monitor the adoption of e-commerce solutions by conducting periodic surveys to businesses. However, the information about e-commerce adoption is often available online in each company corporate website, which is usually public and suitable for being automatically retrieved and processed.

In this context, this work proposes and develops an intelligent system for automatically detecting and monitoring e-commerce availability by analyzing data retrieved from corporate websites. This system combines web scraping techniques with some learning methods for Big Data, and has been evaluated with a data set consisting of 426 corporate websites of manufacturing firms based in France and Spain.

Results show that the proposed model reaches a classification precision of about 85% in the test set. A more detailed analysis evidences that websites with e-commerce tend to include some specific keywords and have a private area. Our proposal opens up the opportunity to monitor e-commerce adoption at a large scale, with highly granular information that otherwise would have required every firm to complete a survey.

Keywords: e-commerce indicator; Big Data; web scraping
