

Data Scraping for Brand Personality Attributes: Five “Free” Text Analytics Tools to Use in the Classroom

Matt Hettche, Christopher Newport University; Michael J Clayton, American University

Unstructured data presents opportunities and challenges for practitioners of direct and interactive marketing. Whereas customer comments from social media sites may prove insightful for CRM and brand development, they are often difficult to synthesize and interpret given their volume and asynchronous nature. Furthermore, conventional methods of content analysis, while applicable to the digital medium, are often slow and inefficient when compared to new automated text analytics platforms. The focus of the proposed research presentation is to evaluate and review some of the “free” text analytics tools that are currently available to students and academic researchers. Our goal is to highlight both the capabilities and usability of various analytics platforms when analyzing for brand personality attributes within customer comments (a relatively routine and useful application for unstructured data). Following Dean 2014, we evaluate the functionality of text analytics tools on two basic criteria: information retrieval and/or content categorization. Emphasis during the presentation will be placed on comparing the commercial “freemium” platforms with other available open source developer tools. The ultimate goal of the proposed research is to showcase current text analytics platforms that can be integrated into advanced undergraduate and/or master’s level marketing courses.

To contextualize the issues and capabilities of the available tools, we plan to analyze the customer comments from the Facebook fan page of the Dollar Shave Club. The Dollar Shave Club is a (relatively) new subscription based product that relies heavily on referral and customer word-of-mouth. Using a framework by Aaker 1997, we aim to demonstrate how the customer comments of the Dollar Shave Club can be analyzed with five free text analytics platforms for its own unique brand personality. In addition to commenting on each tool’s strengths and weaknesses, a brief screencast of each tool’s demo/overview will be made available.

Platforms/tools to be evaluated:
Data Extraction

MS Excel VBA (developer tools)

There are two basic methods of gathering unstructured data with Excel (1) extracting a table of data from a website using a VBA query and (2) scraping a website’s HTML in VBA and parsing its HTML. Data can be extracted and imported into an Excel Spreadsheet or a CSV file.

iMacros (free add-on for the Chrome and Firefox browsers) [similar to WebScrapers, Parsehub, and DataMiner). iMacros is a simple data mining extension for facilitating online research when the goal is to get unstructured data into a spreadsheet. It is intended as an easy-to-use tool for beginner and intermediate users. It has a premium or full “pay for use” version.

Python & Aptana Studio 3 (open source developer software)

Python is a programming language often utilized by practitioners. It is open source and has advanced features for abstracting data; Aptana Studio 3 is a free (open source) Python editor that can be used to execute various data extraction commands. Simple examples will be provided.

Content Categorization

SPSS Modeler (free with an IBM Academic Initiative membership)

A visual interface allows data content structuring according to various pre-defined data

extraction rules. Techniques range from beginner to advanced.

Full-Suite Commercial Tools

Simply Measured (one free social media report for Twitter, Facebook, Instagram, Google+ and Vine). Similar to Simply Measured are the commercial sites Mozenda and Automation Anywhere (each come with free 30 day trial).

Bibliography

Aaker, J. (1997). Dimensions of brand personality. *Journal of Marketing Research*, 34(3), 347-356.

Dean, J. (2014). *Big Data, Data Mining and Machine Learning: Value Creation for Business Leaders and Practitioners*. Hoboken, NJ: Wiley.

Neuendorf, K. (2002). *The Content Analysis Guidebook*. Thousand Oaks, CA: Sage Publications.