

Applied Webscraping in Market Research

Herrmann, Markus^a and Hoyden, Laura^b

^aData Lab, Data & Technology, GfK SE, Nuremberg, Germany

^bMarketing & Data Sciences, Data & Technology, GfK SE, Nuremberg, Germany

Abstract

Modern Webscraping tools and APIs facilitate the extraction of information from the Internet significantly. We outline, that Webscraping, as a common practice to load, prepare and statistically analyze specific structured or unstructured data from the Internet, has become an essential technique in Marketing and Data Science. Furthermore, we emphasize the importance of Open Data and social media data as a scraping target. While we argue that Webscraping of internet data is an enabler and driver of product innovation in Market Research, it should also be noted that just gathering and integrating more data cannot replace research and modeling expertise; and that focusing on easily available data only, may inevitably lead to wrong conclusions or cause legal issues in commercial environments. As an result, data management concepts have to be applied to ensure accuracy, comparability, findability, re-usability and legality of the scraped data. In this presentation we discuss how data lakes, (meta-)data management and data integration processes help to extract most insight of scraped data.

Keywords: *Market Research, Marketing Science, Data Science, Webscraping, Open Data*
