Web Analytics Tools: Towards Taxonomy-Based Archetypes

Masterarbeit

zur Erlangung des akademischen Grades „Master of Science (M.Sc.)“ im Studiengang Wirtschaftswissenschaften der Wirtschaftswissenschaftlichen Fakultät der Leibniz Universität Hannover

vorgelegt von:

Name: Okpure Vorname: Joel

Geb. am: in:

Prüfer: Prof. Dr. M.H. Breitner

, den 25.10.2023

Table of Contents

1 Introduction

Resulting from its growing impact on individual consumers and major economies, the internet has evolved into an essential component of the daily existence (Petrosyan, 2023b). With a total of 5.3 billion users in 2022, which represents 66% of the global population, the number of internet users worldwide was estimated to increase by 0.4 billion compared to 2021. This is due to easier access to computers, the modernization of countries and increased utilization of smartphones allowing people to use the internet more frequently and conveniently (Petrosyan, 2023a). In the past, companies invested in their websites primarily to maintain their brand value or to keep up with their competitors (Järvinen and Karjaluoto, 2015, p.1). However, as digital interactions become more prevalent, customers are increasingly engaging with businesses through digital channels (Singal et al., 2014, p. 24). This has led to a heavy reliance on web channels for product commercialization. Marketers have therefore come to recognize the importance of monitoring online interactions and measure the performance of their channels. Hence, enterprises operating in the online world invest billions of dollars in Web Analytics (WA) to generate significant online revenue (Lovett et al., 2009, p. 1; Järvinen and Karjaluoto, 2015, p. 1). The Web Analytics Association defines WA as the systematic process of "objective tracking, collection, measurement, reporting, and analysis of quantitative internet data to optimize websites and web marketing initiatives" (Singal et al., 2014, p. 24).

Today, the adoption of WA, which is a crucial step towards quantifiable marketing, has become the primary source of revenue for nonhuman sales efforts (Järvinen and Karjaluoto, 2015, p.1). In 2014, more than 60% of the top ten million most popular websites around the globe use WA (Järvinen and Karjaluoto, 2015, p.1). Nevertheless, the evidence regarding the benefits of using WA for digital marketing performance measurement is contradictory in academic research. On the one hand, some case studies have demonstrated that using WA to measure and optimize digital marketing performance can improve the efficiency of marketing actions and increase sales revenue. On the other hand, WA is only used on an ad-hoc basis, and the metrics data are not utilized for strategic purposes, making it unclear what benefits it provides (Saura et al., 2017, p. 1f.; Chaffey and Patron, 2012, p. 30). Added to this are the growing challenges posed by the digital transformation. Customer engagement is evolving rapidly in the digital realm, driven by rising expectations for satisfaction in their moment of need (McCormick et al., 2017b, p. 2). According to McCormick et al. (2014), traditional WA methods were not originally created to handle the wide range of channels, devices, and the rapid pace at which modern digital interactions occur today. However, although WA has been in existence since the inception of websites, it remains in a state of constant evolution. Therefore, WA is continually adapting to the shifting landscape of digital customer engagement and the digital transformation of businesses (McCormick et al., 2017a, p. 2).

To be able to utilize WA, having an appropriate WA system is crucial. As WA developed, so did the tools and services. There is a wide range of WA products and services available in the market, varying in sophistication and form. The range of options spans from free basic solutions to high-end products costing several hundred thousand euros. Selecting a WA tool has significant long-term strategic implications since these tools serve as the basis for making competitive decisions, and the quality of the analytics directly affects the effectiveness of those decisions. Therefore, finding the right product that fits one's specific needs within this spectrum can be quite challenging (Nakatani and Chuang, 2011, p. 172; Hassler, 2009, p. 37).

Although WA is extensively utilized by well-known websites, its increasing popularity among users is not reflected in the academic research (Saura et al., 2017, p. 1). Thus, the objective of this research paper is to provide a holistic overview of available WA tools with their corresponding functionalities to simplify the tool selection process. This aim will be approached by the following research questions:

RQ1: What elements does a taxonomy of Web Analytics tools consist of?
RQ2: What archetypes of Web Analytics tools can be distinguished, based on a cluster analysis?

Regarding the procedure, chapter 2 starts by providing a theoretical foundation, which consists of the relevance of WA in general, followed by an introductory description of a WA system, finalized by a short description of the WA market. In chapter 3 the research design is thoroughly described, whereas chapter 4 is dedicated to the full process of the taxonomy development and application. This starts off with the iteration circle, followed by the evaluation and the final taxonomy. After that, the taxonomy is used to perform a cluster analysis with the goal of the identification of the archetypes in chapter 5. In chapter 6, the decision tree building is approached, which is used for prediction and classification purposes. Subsequently, the discussion of the given results and further implications are made in chapter 7 and after the presentation of limitations that occurred during the research in chapter 8, chapter 9 rounds up the research paper with a conclusion.

9 Conclusion

Since the foundational framework of WA is inherently insufficient in addressing the growing array of communication channels, sophisticated consumers, technical complexities, and the widespread adoption of analytics across data-driven organizations, WA had to evolve. To adapt their WA solutions to modern conditions, certain vendors have dedicated themselves to DI practices. These vendors offering modern WA technologies are not only part of a broader DI landscape but are also amongst the leaders of the entire WA market (McCormick et al., 2014).

At present, despite the ongoing growth in the use of WA tools, there is no comprehensive view or categorization of the extensive selection of available WA tools, along with their distinct functionalities and attributes. With the development of a taxonomy of WA tools, this research paper can be viewed as a starting point in this regard. The aim is to provide a holistic view of the main functionalities within WA systems.

Starting with the initial identification of relevant literature and real-world WA tools, the basis for the taxonomy development was created. In an iterative procedure, the resulting analysis of literature and WA tools successively expanded the taxonomy further, until an evaluation by practitioners finalized the taxonomy. Thereby, five expert interviews were conducted, in which the completeness and usefulness of the taxonomy was ensured. The final taxonomy consists of six perspectives and 30 dimensions with a total of 106 characteristics, which are collective exhaustive and mutually exclusive. The basic taxonomic structure is thereby divided a functional part containing the four perspectives *"data management and availability"*, *"reporting and analysis functionality"*, *"integration support"*, and "*services and support"* and a non-functional part containing the two perspectives *"administrative control"* and *"legal management"*.

Based on an internet search of 61 identified WA tools, a distribution was created to assign each tool a characteristic from each dimension. The result allowed for an initial comparison of the distribution of certain functionalities. Subsequently, a cluster analysis was used to identify three different archetypes of WA tools. Archetype 1 was named "web oriented, standardized tools based on client-side data collection", archetype 2 was named "high-end digital intelligence tools", whereas archetype 3 was named "diverse and customizable tools with legal orientation". Archetype 1 consists of 28 tools and represents the more traditional part of WA, in which most data is captured from websites or web apps and mostly standard functionalities are offered, whereas archetype 3 consist of 30 tools and represents a diverse set of WA tools with different kinds of features that are also customizable in most cases. The most striking one is archetype 2, which consists just of three tools. As the name implies, this archetype represents the modernized part of WA with a high number of functionalities used in the DI area. The cluster analysis indicates that there are a lot of use

cases in WA and there is a tool that offers the corresponding set of functionalities tailored to the required use case. Another indication is that there are lots of dependencies between the dimensions and characteristics.

Finally, with the use of the ID3 algorithm and the C4.5 algorithm, two decision trees were constructed. Since a decision tree can be used for classification and prediction, the aim was to construct a simple decision tree based on the taxonomy. The resulting decision tree and DR can be applied in practice for example as a guideline for the tool selection process. Therefore, a data set containing 30 attributes, one variable "Tool name" and one target variable "Archetype" with three possible values {one; two; three} were separated into train data and test data, whereby the train data set contained two thirds of the original data set and the test contained one third of the original data set. After both trees were finalized, they were applied on the test data to compare their performances on unknown data sets. Since the decision tree with the higher accuracy was the one constructed with the use of the ID3 algorithm, it was used as basis to derive ten DR. It can be further deduced from the decision trees, that the attributes "data type" and "data warehouse" have the most information gains as the root nodes of both decision trees. Thus, according to the decision trees, they have the greatest influence on the classification of the archetypes. This is a confirmation of previous results because they were among the crucial factors identified in the cluster analysis and play important roles in the naming of the archetypes.

To make a concluding remark, it can be said that WA is currently in an important transition phase towards its modernization regarding DI functionalities. Due to the digital evolution leading to a more and more data driven world, more vendors will think about offering new functionalities to stay relevant. At the same time, the old functionalities will not disappear, which will lead to a market that is increasingly segmented and thus the tool selection will be more difficult.