RESEARCHING MARKETING DATA IN THE DIGITAL ENVIRONMENT TO IMPROVE THE COMPANY'S CONTENT STRATEGY

Digitalization of all spheres of consumer life requires proportionate responses from business. The more stages of purchasing decisions are realized in the digital environment, the more critical it becomes for companies to develop a full-fledged strategy for representation in the digital environment. The first stage of developing such a strategy should inevitably be an analysis of the current position of the company in the digital environment, which requires a new methodology of marketing research based on new methods of collecting, processing and analyzing marketing information. "Old", traditional methods of marketing research, such as empirical research, analysis of CRM systems, etc. are poorly suited to solving this problem.

Fig. 1 shows the author's classification of possible sources of marketing data in the digital environment.

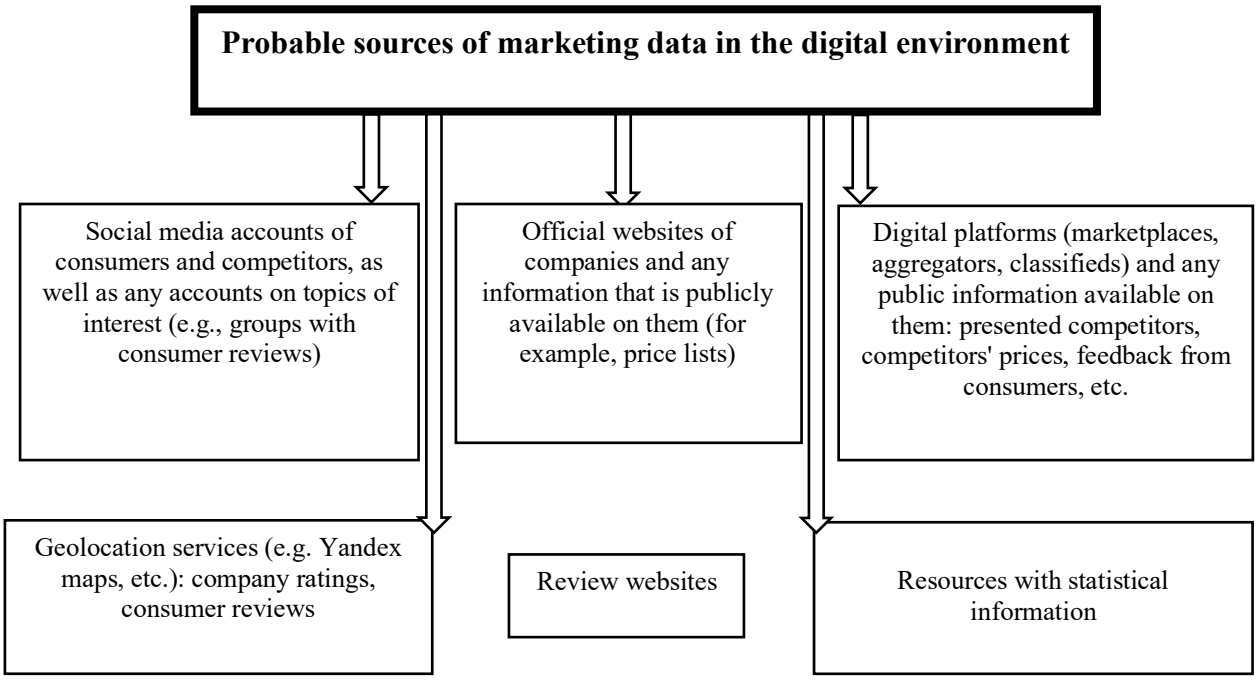

Fig. 1. Possible sources of marketing data in the digital environment
Source: compiled by the authors

To extract marketing data of this type (see fig.1), it is necessary to use a new method that is absent in the classical methodology of marketing research – the method of parsing the digital environment. **Parsing** is the process of automated information extraction from the digital environment based on code written in one of the programming languages (e.g., Python).

Marketing data in the digital environment can be classified in terms of the degree of their structure (see Fig. 2).

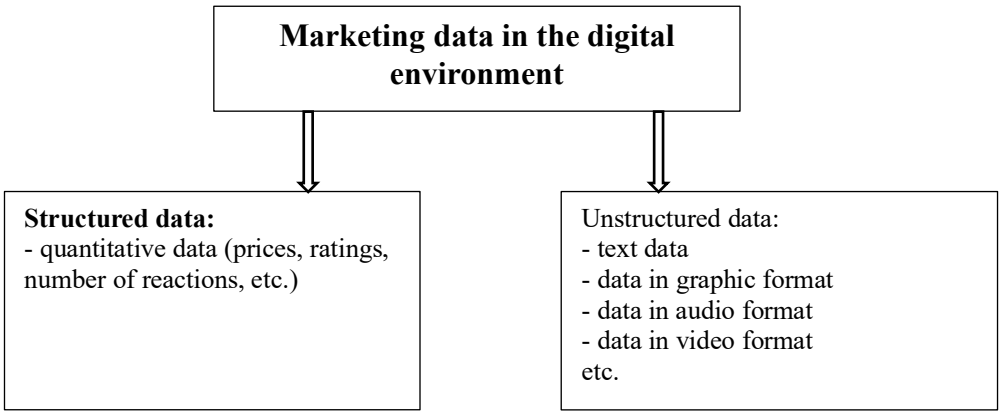

Fig. 2. Classification of marketing data in the digital environment
Source: compiled by the authors

As a rule, unstructured marketing data in the digital environment is presented in the form of natural digital information (NDI). **Natural digital information** is an unstructured array of data that is the result of user engagement in interaction with the digital environment. In this study, we are interested in that part of natural digital information that concerns user interaction with the company in the digital environment, e.g.:

• consumer reviews of goods / services (in cards on marketplaces, in geolocation services, on review sites, in social media, etc.);

• user comments on company publications in social media;

• discussions, debates on thematic forums related to the brand, as well as responses to comments from other users, etc.

The key feature of NDI is that it is publicly available and can be freely researched.

As for the further processing of parsed data, the methods of their analysis are logically divided in terms of the type of data: qualitative, quantitative or mixed. Quantitative, e.g., statistical data on the state of markets and industries, competitors' prices on marketplaces and their dynamics, product ratings, parameters of the target audience's engagement in interaction with the brand, etc. Methods of processing and analyzing such data belong to the field of data mining and depend on their volume, variety, veracity and velocity [3], that is, in essence, whether such parsed data belong to the so-called big data and, if so, to which of its components (Fast data, Big analytics, Deep insight, etc.) [2].

However, of greatest interest in the context of this study are qualitative data extracted from the digital environment, presented in text, audio, video and other similar formats. Methods for processing and analyzing such data lie in the field of computational linguistics, image analysis, and similar areas, i.e. they complement traditional methods of data mining of quantitative marketing data. Popular methods for processing natural digital information include clustering of text arrays, searching for patterns, analyzing anomalies and identifying outliers in data, assessing the thematic, emotional, and contextual components of the NDI. It is also worth noting the role of the neural network approach in analyzing natural digital information, in particular, such methods as automated search for thematic content, analysis of the probability density of the presence of thematic signals [1], analysis of the tonality of reactions of target audiences, methods of semantic modeling and vector representation of words [4], compiling graphs of thematic connections and mutual influences of text information and assessing their structural complexity [5], etc.

Research of the digital environment of an enterprise can be used in making and implementing marketing decisions, e.g., generating demand through the formation of information agenda, adapting the product range and developing packaging, improving positioning strategies, content strategies and community management strategies, setting up tools for attracting paid traffic, etc.

Within the framework of this paper, we would like to demonstrate an example of a methodology for studying the digital environment in order to improve the company's content strategy in social media (see Fig. 3).

Fig. 3. Methodology for studying the digital environment in order to improve the company's content strategy in social media
Source: compiled by the authors

The digital environment research methodology presented in Fig. 3 can be used to identify anomalies in various data sources (which can, for example, help identify "purchased" reviews and "fake" competitor posts), to compile graphs of thematic connections between different data sources, to compile a conversational dictionary of consumers to deposition from competitors in content in social media, which will ultimately represent a full-fledged strategy for improving (correcting) the company's content in social media as a reaction to the change in the external environment.

Based on the results of the already completed adjustment of the company's content strategy, it makes sense to analyze the dynamics of the seeding matrix. This step will lead to the automation of the presented methodology, since the analysis of changes in the target audience's reactions to the adjusted content strategy of the company can launch a cycle of its automatic adjustment, also implemented through the Python programming code.

The development of this methodology lies in the area of adding other sources of marketing data, in particular, those related to the target audience, positioning, tone-of-voice and other brand identifiers, which can ultimately create a full-fledged, fully automated tool for analyzing the company's current content strategy and improving it based on extracting (parsing) marketing data from the digital environment.

List of sources:

1. Chen, J., Zhang, A., Li, M., Smola, A., Yang, D. (2023) A cheaper and better diffusion language model with soft-masked noise - arXiv preprint arXiv:2304.04746.
2. Hinchcliffe, D. Big Data, The Moving Parts: Fast Data, Big Analytics, and Deep Insight. Available at: https://www.flickr.com/photos/dionh/7550578346/in/photostream/ (accessed 20.01.2025)
3. Kitchin, R., McArdle, G. (2016). What makes Big Data, Big Data? Exploring the ontological characteristics of 26 datasets. Big Data & Society. 3 (1): 205395171663113.
4. Knott, A., Pedreschi, D., Jitsuzumi, T., Leavy, S., Eyers, D. (2024) AI content detection in the emerging information ecosystem: new obligations for media and tech companies. Ethics and information technology.
5. Tufte, E.R. (1998) Envisioning Information. Graphics Press, Cheshire, Connecticut. Sixth printing.