

COMPUTATIONAL ASSISTANCE: newsbot development for financial organizations

ASSESSORIA COMPUTACIONAL: desenvolvimento de *newsbot* para organizações financeiras

ASESORÍA COMPUTACIONAL: desarrollo de *newsbot* para organizaciones financieras

Alberto Marques

Professor Ph.D. of the Graduate Program: Professional Master Degree in Innovation, Communication and Creative Economy at the Catholic University of Brasília (UCB). Bachelor in Journalism.


alberto.marques@gmail.com

 0000-0003-1861-2630

Suélem S. P. Barroso

Master in Innovation, Communication and Creative Economy (UCB). Bachelor in Social Communication – Journalism (UNIP).

barroso.sue@gmail.com

 0000-0002-7211-9313

Correspondence: Catholic University of Brasília, Department of Communication, QS 07 – Lote 01 - EPCT – Taguatinga, Brasília/DF – CEP: 71966-700.

Received: 02.16.2021.

Accepted: 04.16.2021.

Published: 07.01.2021.

ABSTRACT:

This research aims to develop the Minimum Viable Product (MVP) of a *cloud* platform that transforms structured data into automated text, to enhance the production of content in Brazilian stock brokerage companies. It is based on applied research that combines different techniques for the development of MVP. As a theoretical path to support product development, it permeates strategic communication in corporations, as well as the adoption of algorithms and content automation in journalistic production. Among the results achieved, the work presents the logic of infrastructure of macro operation for the *newsbot* and the main elements for the many layers of development that an application of this nature needs.

KEYWORDS: Strategic communication; Organizational strategy; Automated content; Newsbot; Algorithms.

Introduction

The sociocultural transformations introduced by information and communication technologies are used as support for thinking about computer-mediated interactive strategies in contemporary organizations. This scenario involves the use of new angles of analysis, with theoretical and methodological implications for organizational communication (BARICHELLO, 2009).

The intensity and speed of the advances already achieved in the communication's field mark the current society in a striking way. On the one hand we see changes in vehicles, processes, work routines and market status related to the various means of communication. On the other hand, the needs for dialogue between companies, public agencies, institutions, personalities and their different audiences change, as well as their demands in relation to the other. To maintain itself as a useful and competent activity, it is necessary to keep up with such advances, adapting to the



new demands and needs of vehicles and advisees, which equally demand increasing and varied skills from communication advisers (FERRARETTO, 2009).

Thus, it is observed that some newsrooms in newspapers and news agencies around the world have been using, for some time, textual production automation and Artificial Intelligence to perform a series of routine tasks (DANS, 2019).

In the last 15 years approximately, algorithms gained ground and established themselves in social scientific work (BEER, 2016). With their help, various routine actions such as bureaucracy reduction, knowledge production and solutions seeking real problems started to pass through a process we call "datafication" (being transformed into data). The Big Data, as well as the development of algorithms, has increased the ability to work with a large volume of information at considerable speed and variety, thus increasing the effectiveness of predictive and prescriptive actions on the social, economic or political behavior of connected individuals, from sophisticated statistical crossovers (MAYER-CHONBERGER; CUKIER, 2013).

Helbing et al. (2017) argue that the data is considered the oil of the 21st Century and that, in 2016 we produced as much data as in the entire history of humankind through 2015. Understanding the interaction between large-scale data collection, algorithmic analysis, computational practices, and public knowledge production is one of the main methodological and philosophical challenges of our time. (MANOVICH, 2011).

Adobe and Econsultancy (2019) explain: companies that treat their data as a critical and valuable competitive asset, while respecting customer privacy, can use it to boost business strategies through improvement of the client experiences as consumers are becoming more demanding and careful about their autonomy.

In this context, organizations need to create a strategic message design that is increasingly aligned with different types of audience and, at the same time, meet business objectives by communicating these messages with scale, accuracy and speed, increasing the personalization of communications. This means communicating no longer massively, but in a personalized way.

Taking that into account, the key question of this work is: how can a financial organization use strategic message design and algorithmic mediation to manage and use the data generated, received and stored, transforming them into coherent automated text?



To help answer it, the objective of this research is to develop a Minimum Viable Product (MVP) of a web platform that transforms structured data from the Brazilian Stock Exchange into automated financial reports.

The MVP development process has gone back and forth because it is a living system in constant evolution. In this work, it is possible to find the newsbot's macro operating architecture logic – created from scratch – as well as the main elements for the numerous layers of development that an application of this nature needs, which goes far beyond a script written in a programming language like Python.

Strategic communication lenses

In an increasingly complex world, organizations dispute for the attention, admiration, affinity, alignment and loyalty of all types of audiences, be they customers, employees, investors and donors, government officials, leaders of special interest groups and the general public (HALLAHAN et al., 2007).

Hallahan et al. (2007) present two communication models¹. The first is called the communication transmission model, which conceptualizes communication as the unidirectional emission of information. Shannon and Weaver's model is a widely cited unidirectional communication model focused on transmitting signals through a channel with a limited feedback capability (SHANNON; WEAVER, 1949, apud HALLAHAN et al., 2007, p. 20). This approach is like a one-way hand: only the organization says what it wants to communicate and the receiver, who is heard little or nothing, assumes the role of a mere listener, reader, viewer. Here the organization communicates in much the same way without taking into account the different audiences and their individual needs.

The second is an interactive model that argues that communication involves the creation and exchange of meaning between parties in a communicational activity (HALLAHAN et al., 2007). In this model, we can cite Wolton (2011), who states that communicating is negotiating and living together. There is no communication without information, and informing is not communicating. Communication is always more difficult because it imposes the issue of the relationship, that is, the issue of the other, and that there is no information without a communication project; in this case the information is the message, and the communication is connected with the relationship, the alterity and the receiver (WOLTON, 2011).

¹ Although there are other communication models, citing and analyzing all of them is beyond the scope of this work.



The author explains that accelerating the production and transmission of an increasing amount of information is no longer sufficient to create an increase in communication. Receivers negotiate, filter, rank, reject or accept the countless messages received. The receiver is increasingly active to resist the flow of information. Yesterday, communication was transmitted, as human relationships were often hierarchical. Today, it is almost always synonymous to negotiate, as individuals and groups find themselves increasingly on an equal footing (WOLTON, 2011).

In this sense, organizations find themselves in the need to create a strategic message design that is increasingly aligned with the different types of audience they have and that, at the same time, meet the business objectives. The newsbot presented in this research is intended to assist in the strategic production of these messages.

That's why companies need to understand that more touch points and higher message volume do not necessarily translate into greater influence (KOTLER; KARTAJAYA; SETIAWAN, 2017). You need to stand out from the crowd and meaningfully connect with consumers at just a few crucial touch points; in fact, just a single moment of unexpected pleasure with a brand is all it takes to turn a client into a loyal lawyer (KOTLER; KARTAJAYA; SETIAWAN, 2017).

Before going specifically into the production of a strategic message design, it is worth emphasizing that this message is inserted in the context of an institutional strategy, in a careful analysis of the consumer journey in which attention, interest, desire and action are mapped to in order to achieve the organization's business goals. A separate message, detached from any strategy, has little or no effect in various contexts and organizational situations.

This implies that if the MVP of this work is not inserted within an adequate institutional and communication strategy, it will not be an effective tool for the organization, regardless of its nature.

Since the appearance of International Journal of Strategic Communication (IJSC), in 2007, the term "strategic communication" has become popular both in academia and in practice as a solution to many pragmatic problems. The authors list four dimensions for strategic communication (ZERFASS et al., 2018).

First, they present that the term "strategic communication" works as a substitute for "integrated communication", a generic term for all types of goal-oriented communication, initiated by organizations to address any type of interested parties and audiences.



Second, there is a growing debate about the new role that communication is assuming in contemporary large, private and publicly traded companies that need to be strategic and decision-making, as opposed to tactical and supportive. The main contribution of this vision is the notion of organization as a corporate actor, incorporated into society and characterized by interests, structures, processes, cultures and specific modes of decision-making interdependent at other levels (macro and micro).

Third, there is an old but growing interest in communication in the context of military and national power (GRAHAM, 2017; NOTHHAFT; SCHÖLZEL, 2015; PAUL, 2011, apud ZERFASS et al., 2018). Fourth, "strategic communication" is used as an alternative terminology for the established discipline of public relations, without altering the underlying research objects or perspectives of that field (ZERFASS et al., 2018, p. 489).

In order to guide this study, we will continue with the second approach to strategic communication. Hallahan et al. (2007) expose that strategic communication differs from integrated communication because its focus is on how an organization communicates through organizational efforts. The emphasis is on applying the communication strategy and on how an organization works as a social actor to advance in its mission.

An organization's ability to communicate strategically is made up of a multitude of sub-processes that take place between co-workers, managers, boards, CEOs and external stakeholders on a daily basis. Therefore, it is necessary to consider interaction processes not only as important in themselves, but also as part of the strategic communication and overall performance of an organization (HEIDE et al., 2018).

It is important that strategic communication directs the organization as a social actor that acts to promote itself and its products, its causes and the social movements in which it is associated through the intentional actions of its leaders, employees and communication professionals (RUÃO, 2016).

In that regard, a brief understanding of the strategic aspects can, to some extent, shed light on the discussion. This is because, in the field of communication, strategy has often appeared as a driving force to bring organizations closer and closer to their business goals with their audiences.

In this perspective, strategic communication could be better used if it were present in organizations that have a structure as a living system in constant flow and transformation, because organizations of this nature are more susceptible to innovation as they develop the capacity to transform the environment more quickly



and promptly adapt to this change since they have greater flexibility to solve problems in a pragmatic way.

That said, strategy is understood as a social practice in which the act of making a strategy would be oriented towards answering "where", "how" and "by whom" this work is done (SARAIVA et al. 2011). The "where" would answer where it is inserted in space and time. The "how" would focus on the process in which thought takes place, transforming itself into experience. Last but not least, the "by whom" would correspond to the social actors, actors who are sustained by practices to act and whose activities are fundamental for the organization's survival.

Communication is the "glue of the organization", as it allows coordination to be achieved with reduced effort and minimal cost control (RUÃO, 2016).

From this angle, we could say that, in general terms, strategic communication is the confluence of practices, and it would be possible to gather some vital elements for its performance, which includes the understanding of situations where there is no fixed and unshakeable point of view, as well as the establishment of the symbiosis between the institutional strategy and the strategic communication which, altogether, help the organization or a particular product in development to locate itself in time and space so as to achieve its business goals, whether from an institutional or commercial point of view.

In this way, strategic communication could be seen as a hub that connects the institution's business objectives with the internal and external environments, considering relevant non-linear aspects of communication management, such as analysis, planning, execution and evaluation.

For the analysis, the following could be incorporated: benchmarking, in order to learn and get to know about best practices in other companies; the study of stakeholders in order to map and direct communication activities and scenario techniques used to think in different contexts in an effort to prepare the company for the future (VOLK et al., 2017).

The planning, on the other hand, would contemplate the mission and vision of the corporate strategy; communication scorecards, used to translate strategy into action; reporting on results, combining financials and non-financial metrics to show how communications create value; in addition to positioning the brand with the public of interest.

As for execution, managing the relationship with customers, stakeholders, journalists and politicians is fundamental. It deals with change communication



management, which is a structured approach to preparing and helping members of the organization to successfully adapt to organizational change (VOLK et al., 2017).

The fourth element, which corresponds to the assessment, focuses on integrated reports that explain how an organization creates value on six different aspects: financial, commercial, manufacturing, intellectual, human, social/relationship and natural capital; it also includes a dashboard in which essential information is presented, as well as the evaluation of the process that focuses in the development, progress and constant search for improving the project

The four macro elements mentioned comprise some activities performed by members of an organization in the areas of business, operations, communication and marketing. This indicates that strategic communication is transversal when moving through different sectors of the organization. The act of strategizing is not a manual or something rigid.

Automated content production and journalism

In communication, journalism is one of the areas that most developed writing mediated by algorithms. Computational journalism is a confluence of journalistic practices and knowledge, traditionally associated with Computer Science, in order to develop stories and news based on the social goals and values of journalism in an environment of technological convergence (AZZELLINI; PESCHANSKI, 2018). Diakopoulos (2011) claims that computational journalism is the application of computation and computational thinking to journalism activities, including information gathering, organization and production of meaning, in addition to the dissemination and public response to news information; all this while maintaining the core values of journalism, such as accuracy and verifiability.

It should be said that computational thinking is understood as the ability to navigate through different levels of abstraction (AZZELLINI; PESCHANSKI, 2018). In journalism, for example, the computational way of thinking helps to understand how a certain technology works and what possible professional appropriations can be made from it.

In addition, it helps to understand the innovation processes triggered by its structure. It may be precisely through the acquisition of new knowledge that the bridge is established between the functions of journalism, in today's society, and the ability to obtain socially relevant information through computer systems (LIMA JÚNIOR, 2012 apud AZZELLINI; PESCHANSKI, 2018).



One of the specificities that characterize computer journalism is that it is based on structured narratives, that is, narratives developed from structured data with degrees of organization, systematized and arranged by standards of values. It is in this horizon that automated journalism emerges, a relatively new phenomenon in the field of computer journalism. Terms such as robotic journalism, automated journalism, algorithmic journalism, or machine-written journalism dominate the media and scientific discourse (DÖRR, 2016).

Automated journalism, is the term that will be used in this work, which uses Natural Language Generation (NLG), a technology defined as software and computer systems that automatically produce the representation of human language (natural) from a computational representation of information (REITER; DALE, 2000).

Automated journalism refers to the process of using software or algorithms to automatically generate new stories without human intervention, after an initial programming of algorithms (GRAEFE, 2016, p. 14). Therefore, once an algorithm is developed, it allows you to automate every step in the process from producing a news story, from collecting and analyzing data, to creating and publishing it. This method works with fact-based stories when clean, structured, and reliable data are available. In every situation, algorithms can create large-scale content, customizing it to each reader's individual needs, faster, at a lower cost, and potentially with fewer errors than any human journalist (GRAEFE, 2016, p. 14).

The automation can be seen as part of a larger and longer epistemological transformation in journalism, known as the "quantitative turn" (CODDINGTON, 2015), rearticulating values and practices of apprehension and representation of social reality and making the rationalities of journalism increasingly dependent on quantification and computation (MILOSAVLJEVIĆ; VOBIČ, 2019). It is always worth remembering that, although we are talking about journalism, theoretically it would be possible to employ the ideas discussed so far in various segments of modern social life.

Automated journalism is defined as autonomous content production through computer algorithms (WÖLKER; POWELL, 2018). The concept can also be understood as a set of computational skills needed to collect, process, combine and visualize data in order to produce news. Finally, complementing everything that has been said so far, Strauss (2018) claims that automated journalism uses a precise set of data to make sense of the information (TABARY; PROVOST; TROTTIER, 2015).

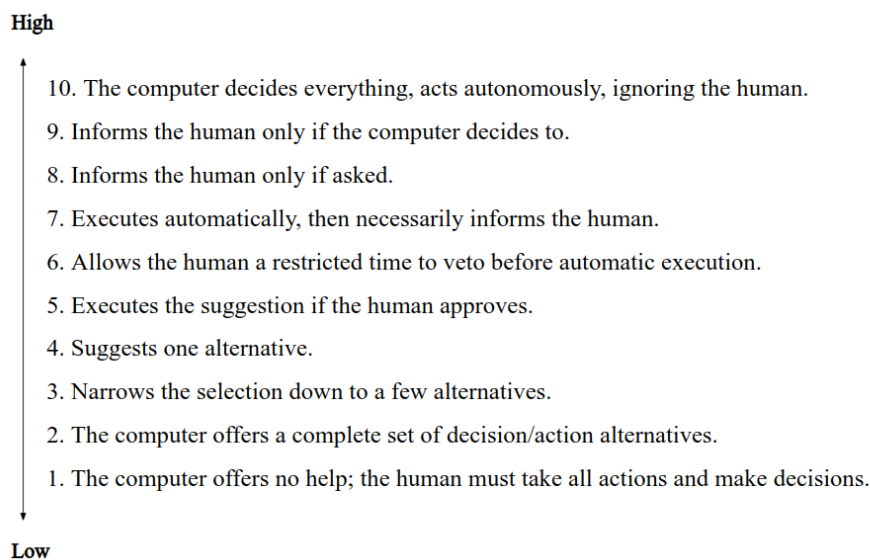
About automation, it is relevant to complement that there are different levels of automation, and understanding these degrees is a fundamental question. Diakopoulos



(2019) explains that automation, in turn, was defined as "a device or system that performs (partially or completely) a function that was previously, or conceivably, performed (partially or completely) by a human operator".

The term Artificial Intelligence can be understood as a computer system "capable of performing tasks that normally require human intelligence". The term "autonomous technology" implies a version of full automation in which a system operates without human intervention; setting aside the design, the maintenance work is also done by humans, a thing that all designed systems need. Full autonomy is one extreme in a spectrum of options that combine humans and computers (DIAKOPOULOS, 2019, p. 16).

Figure 1. Automation levels that combine more or less human effort and automation.



Source: Diakopoulos (2019, p. 17).

By looking at the figure above, it is possible to identify that, in all aspects, the human is a fundamental part, whether for design, complete or partial decision-making, in addition to maintaining the application. After a brief analysis, we could say that automated journalism goes through points seven and eight, and that this hybrid human-machine relationship does not nullify the human being; on the contrary, it can enhance their creative and critical capacity, thus providing deeper analysis, more



elaborate interviews, more refined perceptions of the agenda and more assertive strategies.

Current solutions range from simple code that extracts numbers from a database, such as those used to fill in the blanks in pre-written story templates, to more sophisticated approaches that analyze data in order to obtain additional information and create more attractive narratives (GRAEFE, 2016).

Figure 2. How algorithms generate news.

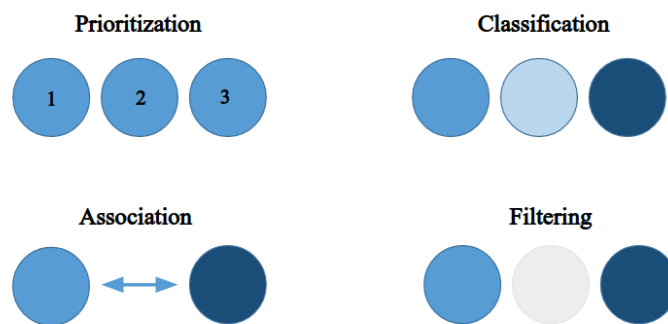


Source: Graefe (2016).

Machine learning algorithms make decisions based on data and there are at least four decision making that algorithms do: prioritization, classification, association and filter, which can be composed of several other sub-decisions (DIAKOPOULOS, 2019). For example, to obtain a news summary, each algorithm must first filter or select a subset of sentences representative of the news and then prioritize them in terms of importance to the user, before presenting them in the summary. Other composition of decisions are also possible.

Figure 3. Diagram of the four fundamental algorithmic decisions.





Source:

Diakopoulos (2019, p. 21).

In summary, algorithms can perform cognitive tasks by means of equation-based calculations or transforming digital bits into words and symbols very quickly, as well as making decisions related to prioritization, classification, association, filtering and compositions based on the basic decisions (DIAKOPOULOS, 2019).

By understanding this logic in which news is generated automatically and basic decisions are taken practically autonomously, the possibilities are abundant.

Research method and techniques

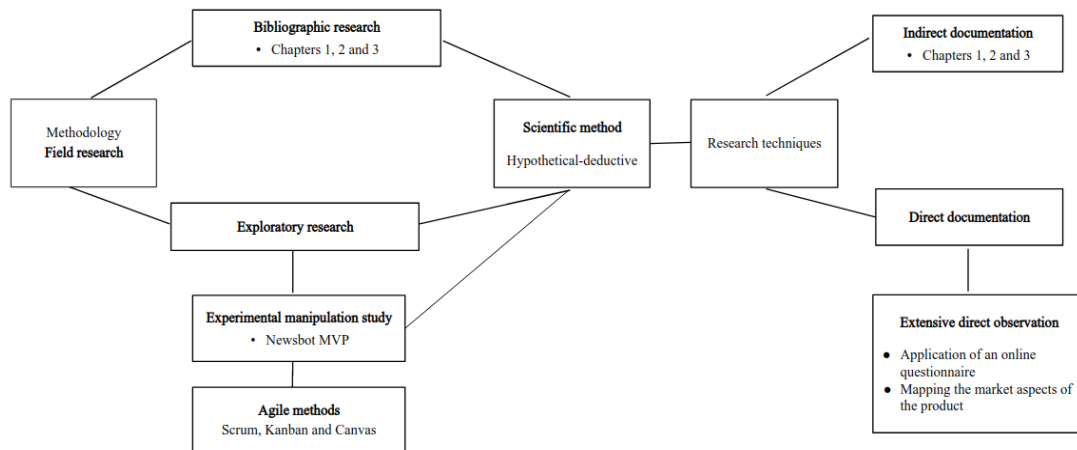
The scientific method of this research is the hypothetical-deductive. And if the method is the path (LAKATOS, 2010), we could then say that research techniques would be the way to walk the path.

In that regard, Lakatos (2010) explains that research techniques are divided into two spheres: indirect documentation, covering documentary and bibliographic research; and direct documentation, which is subdivided into intensive direct observation (observation, interview) and extensive direct observation (questionnaire, form, opinion and attitude measures, tests, sociometry, content analysis, life history, market research).

We use indirect documentation as a research technique for the bibliographic dimension; direct documentation, in the aspect of extensive direct observation, with the application of an online questionnaire (answered in writing and without the presence of the researcher (LAKATOS, 2010); with the stock brokerage companies XP Investimentos, Guide Investimentos and Terra Investimentos to map aspects that could guide the design of the MVP. The methodological aspects of this research are summarized in the following picture:



Figure 4. Outline of research methodology.



Source: the authors.

MVP development processes

Sommerville (2007) explains that new software must be developed quickly to take advantage of new opportunities and respond to competitive pressures in the market and that, many times, they are not geared towards speed in their development. Often fast software development processes are designed to create a useful application quickly, being made available in a series of increments, each of which includes new system functionality. Therefore, the MVP will introduce the increments of the project's primary need: writing automated text from structured data.

In the 1980s/1990s, it was understood that the best way to get good software was through careful project planning. The dissatisfaction with this type of approach allowed the emergence of agile methods to deliver a working software quickly to customers, with changes for later interactions of the system (SOMMERVILLE, 2007).



Extreme Programming, Scrum, Crystal and Kanban are some types of agile methods applicable to software development. Although all of these are based on the notion of development and incremental delivery, each one of them proposes different processes for achieving this (SOMMERVILLE, 2007). With respect to agile work methods, we reconciled Scrum and Kanban to develop the prototype and Canvas to map the market aspects in which the application could be inserted.

Scrum is an agile method as it follows the principles of the Agile Manifesto² (SOMMERVILLE, 2018). Also considered a framework (CRUZ, 2018), it is used in the fast management for creation and development of projects for any complex product, mainly because it is interactive and incremental, which optimizes predictability and risk control. The main idea of Scrum is to allow a small team of people to handle and solve complex and adaptive problems while creatively delivering products (CRUZ, 2018). In this work, we take Scrum as a framework.

For the implementation of any control of empirical processes in Scrum, three pillars are needed: transparency, inspection and adaptation (CRUZ, 2018). When viewing these three pillars, the sprints start, the moment in which the game actually happens. Cruz (2018) presents four formal events that make up the sprint: planning, meeting, review and retrospective. Within Scrum there is also a division of roles and responsibilities. The teams are small and perform events with a fixed duration. Basically, the Scrum team is composed of three roles: Scrum Master (SM), Product Owner (PO) and the Development Team (TD)(CRUZ, 2018).

The SM is responsible for ensuring that Scrum is understood and applied; it's like the team coach. The PO, the main product backlog manager, guarantees the value of the work performed by the team and, in addition to keeping the product backlog, it makes it visible to everyone. TD, on the other hand, performs the development and transforms the product backlog into increments of functionality, creating a ready-made system that can be delivered to the customer.

The MVP development project of this research initially adopts some elements of Scrum so that, when the application is available to the business public, the framework can be applied in its entirety. Regarding Scrum roles and responsibilities within the project, we have the research authors occupying the SM and PO positions and two programmers as the TDs.

2 The Manifesto for Agile Software Development was created collaboratively in 2001 by 17 professionals representing software development methods.



Another approach used in the development of the prototype, the Kanban, is a manufacturing method oriented towards series production, its development being credited to the Toyota Motor Company. This method is applicable only in discrete and repetitive production systems (LOBO, 2010). We apply it to the online tool Trello to have, in an organized way, the SM/PO workflow with the TDs regarding the sprints.

In relation to MVP, the project assumes the evolutionary model as the software lifecycle of a software process. The life cycle, term used in this work, is a set of activities that lead to the production of a software product (SOMMERVILLE, 2007).

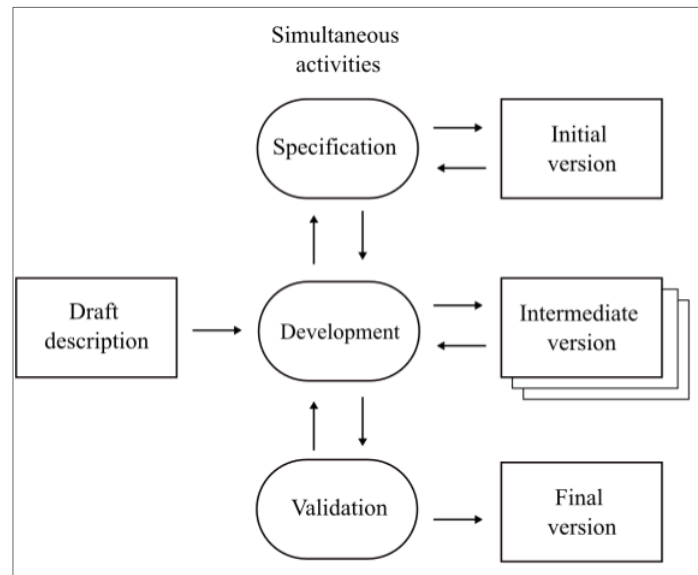
The evolutionary development approach is based on the idea of developing an initial implementation, exposing the result to user feedback and refining that result through multiple versions until a suitable system is developed (SOMMERVILLE, 2007).

Thus, the advantage of a software process based on the evolutionary approach is that the specification can be developed incrementally, as this is the best development method for small and medium-sized systems (up to 500,000 lines of code, in the case of the MVP of this work). As the newsbot project advances and reaches a large dimension, in the future a mixed process can be incorporated, absorbing characteristics of other models.

In this work, we use the throwaway prototyping to solve the uncertainties in the application specification, as shown in the figure below, whose objective is to understand the customer's requirements and, from that, to better develop the definition of requirements for the system. (SOMMERVILLE, 2007).

Figure 5. Evolutionary development model.





Source: Sommerville (2007).

The development of the MVP is divided into three stages: (i) specification and design; (ii) development; and (iii) test (SOMMERVILLE, 2007). In the first, there is no detailed system specification, and the project documentation is minimized or generated automatically, including: global strategic planning of the project, initial design of the macro operation of the software, schedule, definition of servers, hiring of staff and mapping of future application challenges.

The second is the system developed in a series of increments, focusing on the development itself, in which each increment is driven by a sprint of at most two weeks, an element of the Scrum framework, combined with Kanban.

The third step is related to the system's user and interfaces, which generates a web-based interface for a browser. For this, it considers testing cycles, validations, implementation of improvements and metrics. Each increment is designed, developed and tested before the next enhancement specification or new increment.

To map the market aspects of MVP, we use the Business Model, Canvas, a scheme to describe, visualize, evaluate and change business models (OSTERWALDER; PIGNEUR, 2011).

Study of experimental manipulation: the MVP

The product developed by this research is an MVP of a web platform that transforms structured data from the Brazilian Stock Exchange into automated financial



reports, based on dynamic templates, combining human and Artificial Intelligence³ to enhance the production of content in stock brokerage companies that need to write to different types of investors with scale, precision and speed to give flow to the growing volume of information in the financial market.

At first, the newsbot will begin to learn and write about the Brazilian financial market and, later, other segments will be incorporated into the learning process. Automated texts are produced in Portuguese; other languages such as English and Spanish are planned for a more advanced stage of the application.

The development of the platform prototype follows the good practices of non-functional requirements of the ISO/IEC 25010:2011, which determines which quality characteristics to consider when evaluating the properties of a software product.

Regarding the quality model of a system, in principle, the MVP encompasses functional suitability, performance efficiency, security and maintainability. The other features (compatibility, usability, reliability and portability) will be incorporated at a later stage (ISO/IEC, 2011).

Figure 6. Characteristics and sub-characteristics of the product quality model defined in ISO/IEC 25010:2011.



Source: ISO/IEC 25010:2011.

The programming language used for the development of MVP is Python, which is an object-oriented language, which allows data and methods to be easily encapsulated and reused. As a dynamic language, it allows attributes to be added to objects in real time and variables to be dynamically typed, facilitating fast development. It has an extensive library, including components for graphical programming, numerical processing and web connectivity (BIRD; KLEIN; LOPER, 2009).

³ Artificial Intelligence will be incorporated into the MVP at a later stage of the application.

Open source Artificial Intelligence and machine learning libraries (Keras, TensorFlow e Scikit-learn), of analysis and processing of scientific data (NumPy e SciPy); of generic data analysis (Pandas), data visualization (Seaborn) and natural language processing (Natural Language Toolkit (NLTK)) accelerate the solution development process, as they offer a set of ready-made tools and implementations, tested and maintained by the community, eliminating the need to rebuild a new computational framework, which allows to focus on new applications. The libraries in use initially in newsbot development are TensorFlow, Pandas and NLTK.

The Integrated Development Environment (IDE) on which the software is developed is Pycharm. In it, you can write and edit source code, debug and test. To transform the source code into bytecode, the code interpretation process is used – since Python is an interpreted language – ensuring the flexibility of maintenance of the software in production and eliminating the need to recompile the code to work in different environments. The bytecode generator used to allow the software to run is the Python interpreter (a virtual machine that converts source code into bytecode).

The newsbot source code is distributed on two fronts, which we name hard code (source code for the software operation, in Python) and web interface code (code intended for the web interface, in Django – free and open source Python framework Web, which encourages fast development, and a clean, pragmatic interface design) (DJANGO, 2005).

The hard code and web interface code are developed in a local environment (PyCharm). From there, they are uploaded to the project folder on GitHub, which automatically connects, once configured, with Heroku, a cloud server for development and initial testing of the application and database. This environment is for the development team to create, improve and test the code. The approval and production environment servers in the cloud, both for the application and for the database, will be those for Heroku and Google Cloud Platform, respectively.

To guide the front-end development, we created the layout of the main screens of the application prototype web interface: login/password screen, dashboard screen and automated text generation screen.

Figure 7. Illustrative application login and password screen.



The image shows a web browser window displaying the login page for 'sintática.com'. The browser's address bar contains the URL 'http://www.sintatica.com'. The page header features the 'sintática' logo. The main content area contains a login form with two input fields: 'User' and 'Password'. Below the 'Password' field is a red button labeled 'Enter' and a link that reads 'Forgot my password'. The footer of the page includes the text 'sintatica.com', a red 'Support' button, '©2020 Sintática', and links for 'Terms of use', 'Privacy', and 'Cookies'.

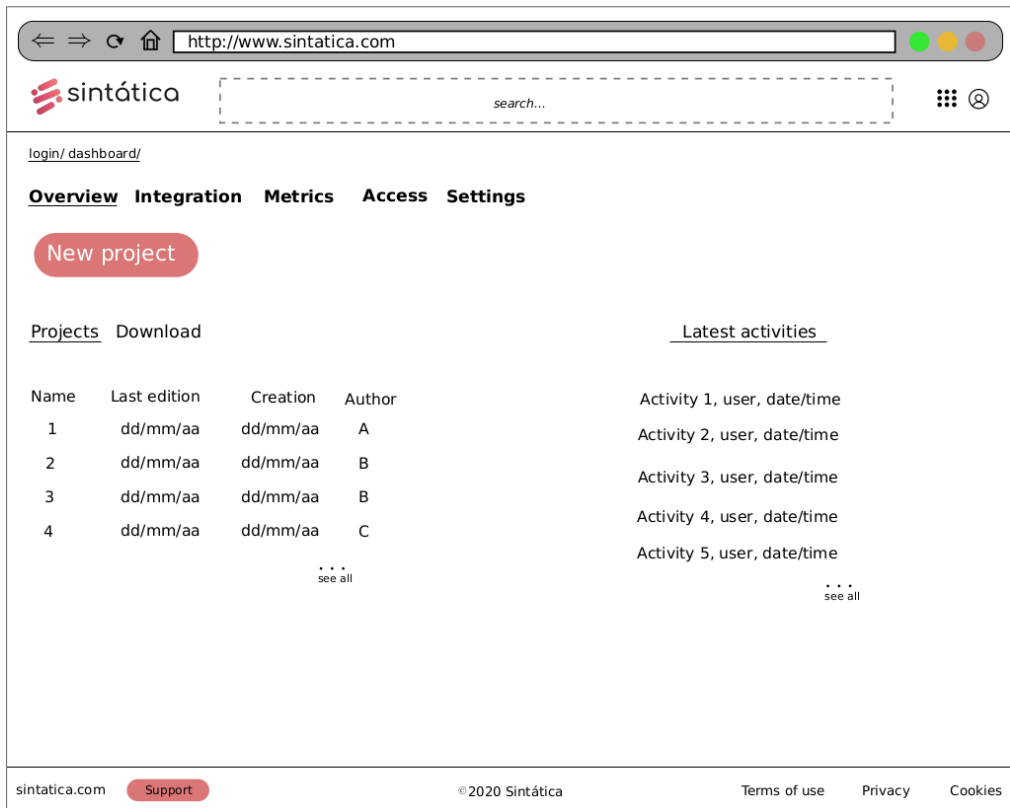
Source: the authors.

The second main screen is the dashboard, in which the user can see the activities performed by whom and when, projects created, integrations with other systems and applications, metrics related to the projects created, access for members of a given team, general settings platform, search tool, tracking of screens accessed, access to user preferences, support, creation of a new project and download of content already created.

At first, the features offered by the dashboard screen for MVP are focused on the Overview tab (Figure 8) and the New Project button (Figure 8), which takes the user to the automated text generation screen. That's because all the efforts in this work are directed towards the development and improvement of the hard code and the necessary integrations between the source codes.

Figure 8. Illustrative application dashboard screen.



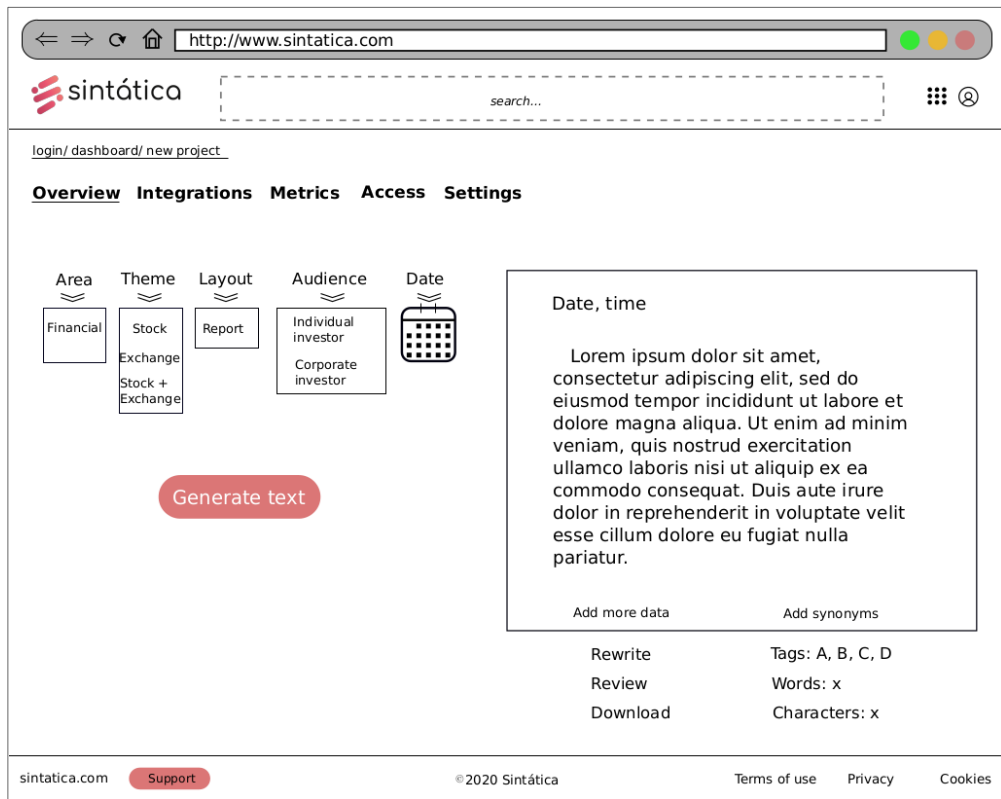


Source: the authors. The screenshot for automated text generation is

where the user can select the filters Area (finance), Theme (sector and subsector of the Brazilian Stock Exchange), Format (report), Audience (individual investor; corporate investor) and Date (calendar) to then generate the automated text, which will appear in the box on the right side of the screen. Functionalities to enter more data, add synonyms, text proofread and download will be added in increments in the future. The only one available for the alpha version of MVP will be Rewrite.

Figure 9. Illustrative application text generation screen.





Source: the authors.

T

that said, at this first moment, we tested the

following macro logic of how MVP works (Figure 10), which presents seven steps for the production of automated text. We use as a basis the five steps presented by Graefe (2016), adapting them to the reality of this project. In what follows, the numbers in parenthesis refer to the 7 steps mentioned above and pictures in figure 10 below.

After the user is connected to the platform with login and password (1), have generated a new project (2), and inserted the filters in the automated text generation screen (3) then, by clicking on the "generate text" button, the user sends the input data for the database server that has the management software (hard code), as well as new, historical and contextualized data (4).

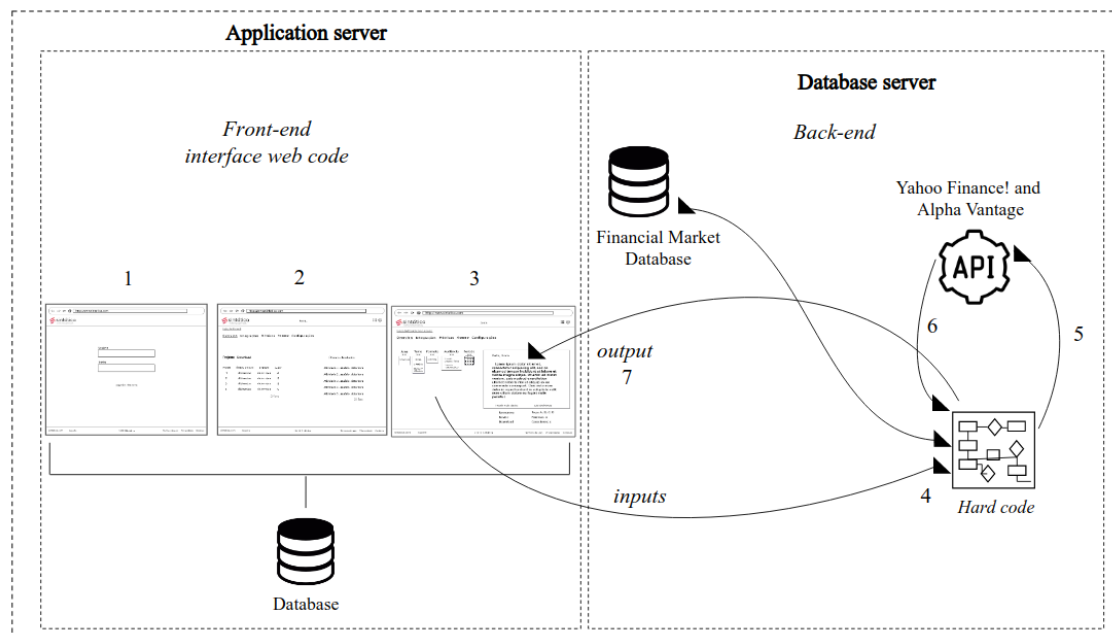
The algorithms then collect/request the Yahoo Finance! and Alpha Vantage APIs: Ibovespa score, foreign currency quote, stock prices and variations, best and worst performance of companies by sector (5), which returns to management software (hard code) with analysis of data correlations and pre-defined rules by topics.

After this step, but still in the management software, the prioritization of insights starts from the predefined rules - where the organization of predefined news elements takes place following rules to generate a narrative where dynamic templates fit (6); lastly we have the output generation (7), that is, the automated text generated and sent to the text box on screen 3 (see figure 9) of the web interface. This generated text can be published automatically, after integration with the user/company's



management and publishing system, or copied by the user and posted/published manually.

Figure 10. Macro logic of the MVP development infrastructure.



Source: The authors and project development team.

The planning and development of the code accounted for around 280 working hours from March/20 to January/21. The MVP, up to the last mentioned date, has about 42 thousand lines of programming.

Memorial: criation process

In the first semester of 2019, we started a deep journey of investigation and understanding of what was being researched, developed and applied in universities, startups, newspapers and news agencies around the world about automated texts.

During that same year, we sought the team to develop the newsbot aimed at non-journalistic organizations. We talked with some developers about the project presentation, product development possibilities, Application Programming Interface (APIs), time, investment and execution process.

The possibilities were clear, which helped us to have a macro view of the project and know if we were going in the right direction. However, no team accepted to develop the software due to unavailability of time.

At this stage, we continued with the research and theoretical foundation and, in parallel, we participated in communication congresses beyond that of Startup Católica



de Brasília, to map the project's economic and financial viability, as well as the premises to compose, to some extent, the business plan of the startup Sintática, a communication technology company that was born in February 2020, in parallel with the academic research of the Master's Program on which this article is based.

This experience allowed us to take a more critical look at some aspects, such as applying an online questionnaire to identify the customer's real needs, the MVP development logic, the attempt to make a pitch with the stopwatch on, besides the administrative, accounting, and legal aspects of a company. The experience was positive and improved our research and our entrepreneurial spirit.

In 2020, we joined the Python Brasil online community to get in touch with developers who use the software development language. The community promptly responded with programming tips, examples and testimonials, allowing us to see how active and collaborative it is. From there, we started conversations with other developers and community members who were willing to support the project and accept the challenge it imposes.

Still in January, Developer A⁴ made the acceptance official, and, on that occasion, some stages were defined, such as specification and design, development and testing, as well as deadlines.

The month of February was devoted to drafting the workings of the newsbot, which would write from structured data, but without templates. After the Master's Qualifying Examination, we reflected on which way to go in terms of technical feasibility and lead time for the MVP development.

The version proposed in the Qualification was disregarded, because it would completely impact the software development, since the programming logic would be different, demanding more time, resources, and staff. When we analyzed the available possibilities, we changed the course of the prototype creation, so that the modeling of the hard code would have the adoption of templates, as is done in most projects.

After that we went back many steps from planning the MVP. We already knew what the new way would be, but we had to redesign how the newsbot would work in this new scenario. This required hours of study and drafts until we got to the first version. The course on programming logic was fundamental, because we literally appropriated computational thinking.

4 The developers will be named by letters that are not associated with their names. We have chosen to mention them like this to avoid any kind of exposure.



As a consequence of the new coronavirus pandemic, the planning and execution of this work was impacted and the schedule was adjusted. The Developer A remained on the project from January to June 2020, and we selected the Developers B and C to continue the project from the point where we had stopped.

In this MVP phase, we incorporated the Scrum framework, which facilitated the definition of sprint tasks, time estimation, management and transparency of the agreed activities, as well as the payment of the hours dedicated to the project, bringing dynamism to the MVP development.

After hiring developers B and C, we found ourselves at the end of September/2020, with the need to look for new team members and started to question where the women were from software engineering, computer science, and mathematics and why we were not finding them.

We then came across several communities that fostered the participation of women in STEM fields, such as the PyLadies Brasil, PretaLab, Women who Code, Rails Girls, CodeGirl, MariaLab among other networks that give voice to women programmers.

We contacted PretaLab, inviting some female developers for a chat. The experience was positive and resulted in the hiring of developer D with a senior profile, who is dedicated to the project 20 hours a week.

In early December, Developer E started on the project, focused on modeling, populating, maintaining the database and connections to the APIs. Dedication of 20 hours a week.

The project followed its course from July 2020 to January 2021, with 12 Sprints in which we had the programming of fundamental increments for the structuring of the prototype both in the hard code, the web interface code and database.

At the end of Sprint12 we achieved two important results: the integration, done on the fourth attempt, of the codes and database; this brought us to the second result: the production of the text itself, including the sectors that are part of the Brazilian Stock Exchange, what was already appearing in the text box of the screen three.

It is worth saying that the access of users, previously registered, has become possible in the development environment and, consequently, the navigation of these users through the software.

To close this section, the stage that refers to the development of the MVP had two challenges. The first was to conceive, in its entirety, the macro logic of the software's operation. It takes many factors, such as diagrams, infrastructure, servers,



integrations, modeling, and staff, which goes beyond a script. Even if the code libraries exist, you still need to know what, when, how and what to use it for.

Carneiro (2016) explains why this knowledge is not made available: the main solutions already implemented in the market come from private Artificial Intelligence companies, which protect their development processes and the teams involved. Today it is possible to access the open code provided by OpenAI.

The second challenge was to acquire and retain specialized staff. Handling this issue is time consuming and requires experience in people management. We realized that, in science, not having a result or seeing that it is completely different from what you imagined is also a result.

Final considerations

This research aims to develop the MVP of a *cloud* platform that transforms structured data into automated text, to enhance the production of content in Brazilian Stock Exchange. To do this, we walked a path where strategic organizational communication and content automation were able to meet.

The understanding of strategic aspects can, to some extent, bring light to this discussion, because in the field of communication strategy has often appeared as a driving force to bring organizations closer and closer to their business purposes with their audiences.

We also approached the need to create a strategic message design increasingly aligned to the different types of audience that such organizations have, because communicating in a massive way today has been less effective, since the receiver chooses and classifies what he wants to consume and, every day, he wants to feel unique, wishing to have a good experience; therefore, the strategic message design is increasingly necessary.

We point out that we do not see algorithms as something evil for any of the actors involved in the process of production or consumption of a text written in an automated way, but the concentration of power and little dialogue can generate and enhance an increasingly distant and distorted relationship between the sender and receiver of the message.

In terms of the experimental manipulation study, we achieved as an initial result the development of the macro logic of the MVP's operating infrastructure. This architecture was fundamental for understanding how the newsbot would work, orchestrating the work of the Development Team and the priorities for the realization



of the project as a whole. The development and incremental testing of the platform's MVP resulted in previously registered users accessing the development environment, enabling navigation through the software and the generation of the first automated financial texts/reports.

As an answer to the question of this research: how can a financial organization take advantage of algorithmic mediation to manage and utilize generated, received, and stored data into coherent automated texts?, we conclude that newsbot can be a mitigating solution to this problem.

However, for the organization to be successful, it is necessary to have a strategic work with the actors involved, so that the product is not seen in the institution as an invader, and the organization's employees do not act as antibodies to expel this as an intruder. The goal is not to replace people, but to establish a hybrid relationship in which humans and machines work together, so that humans have more time to contribute in more strategic and innovative ways.

The hypothetical-deductive method associated with agile methods brought concrete solutions to the research, making us see that nothing is fixed and rigid when it comes to strategy and innovation. It was possible to build probable conclusions from scientific intuition, and the newsbot strategy was not made and followed literally: it was lived daily and redirected to keep the focus on the central objective of the work.

The insertion of a development layer to incorporate artificial neural networks into the newsbot is an increment that is divided into numerous sub-increments. It is like a cake with frosting: to apply this frosting, it is necessary that the whole process: recipe, choice of ingredients, preparation and baking, already has been done. The same is happening with the newsbot, and this has been verified in development. But the entire infrastructure is already being developed to accommodate Artificial Intelligence in the near future, a goal beyond this research.

The newsbot is now part of a larger project, arising from a master's research. It will be constantly improved and more increments will be added to the platform, refining the result through several versions until the system is adequate, as mentioned in the methodology of this work.

To finish, we emphasize that for the newsbot of this research to be a strategic element of an organization and produce messages with strategic design for the investor, it is necessary that it be integrated to the institutional planning, in a way that the stock brokerage company can imprint in its strategy the transforming agent of innovation to aggregate value to the business and to the clients.



References

- ADOBE; ECONSULTANCY. Índice de Experiência: tendências digitais para 2019: Sumário executivo. London: Adobe, 2019. Disponível em: https://www.adobe.com/content/dam/acom/en/modal-offers/econsultancy-digital-trends-2019/pdfs/econsultancy-2019-digital-trends_US.pdf. Acesso em: 3 ago. 2020.
- AZZELLINI, Erica Camillo; PESCHANSKI, João Alexandre. O conceito de Jornalismo Computacional ante as potencialidades das narrativas estruturadas. In: CONGRESSO BRASILEIRO DE CIÊNCIAS DA COMUNICAÇÃO, 41., 2018, Joinville. Anais [...]. São Paulo: Intercom, 2018. p. 1-15.
- BARICHELLO, E. M. M. da R. Apontamentos sobre as estratégias de comunicação mediadas por computador nas organizações contemporâneas. In: KUNSCH, Margarida Maria Krohling (org). Comunicação Organizacional: histórico, fundamentos e processos. Volume 1. São Paulo: Saraiva, 2009.
- BEER, D. The Social Power of Algorithms. *Information, Communication and Society*, [s. l.], v. 20, n. 1, p. 1-13, 2017.
- BIRD, S.; KLEIN, E.; LOPER, E. *Natural Language Processing with Python*. Sebastopol, CA: O'Reilly Media, 2009.
- CARNEIRO, M. S. Comunicação Digital e Jornalismo de Inserção: como o big data, inteligência artificial, realidade aumentada e internet das coisas estão mudando a produção de conteúdo informativo. São Luís: Labcom Digital, 2016.
- CODDINGTON, M. Clarifying Journalism's Quantitative Turn. *Digital Journalism*, [s. l.], v. 3, n. 3, p. 331-348, 2015.
- CRUZ, F. *Scrum e Agile em Projetos: guia completo*. 2. ed. Rio de Janeiro: Brasport Livros e Multimídia, 2018.
- DANS, E. Meet Bertie, Heliograf and Cyborg, the New Journalists on the Block. *Forbes*, [s. l.], 6 Feb. 2019.
- DIAKOPOULOS, N. A Functional Roadmap for Innovation in Computational Journalism. [S. l.: s. n.], 2011.
- DIAKOPOULOS, N. *Automating the News: how algorithms are rewriting the media*. Cambridge, MA: Harvard University Press, 2019.
- DJANGO. *The Web Framework for Perfectionists with Deadlines*. Kansas, EUA: Django, 2005.
- DÖRR, K. N. Mapping the Field of Algorithmic Journalism. *Digital Journalism*, v. 4, n. 6, p. 700-722, 2016.
- FERRARETTO, E. K. *Assessoria de imprensa: teoria e prática*. São Paulo: Summus, 2009.
- GRAEFE, A. *Guide to Automated Journalism*. Columbia Journalism School, 2016.
- HALLAHAN, K.; HOLTZHAUSEN, D.; RULER, B. V.; VERCIC, D.; SRIRAMESH, K. Defining strategic communication. *International Journal of Strategic Communication*, [s. l.], n. 1, p. 3-35, 2007. DOI: 10.1080/15531180701285244.
- HEIDE, M.; PLATEN, S. V.; SIMONSSON, C.; FALKHEIMER, J. Expanding the scope of strategic communication: Towards a holistic understanding of organizational complexity. *International Journal of Strategic Communication*, [s. l.], v. 12, n. (4), p. 452-468, 2018. DOI: 10.1080/1553118X.2018.1456434.
- HELBING, Dirk et al. Will Democracy Survive Big Data and Artificial Intelligence? *Scientific American*, [s. l.], 2017.
- ISO. IEC. *ISO/IEC 25010: System and Software engineering - System and software Quality Requirements and Evaluation (SQuARE) - System and software quality models*. Switzerland: ISO 2011.



- KOTLER, P.; KARTAJAJAYA, H.; SETIAWAN, I. Marketing 4.0. Tradução: Ivo Korytowski. Rio de Janeiro: Sextante, 2017.
- LAKATOS, E. M. Fundamentos de Metodologia Científica. 7. ed. São Paulo: Atlas, 2010.
- LOBO, R. N. Gestão da Qualidade. São Paulo: Érica, 2010.
- MANOVICH, L. Trending: the promises and the challenges of big social data. [S. l.: s. n.], 2011. Disponível em: <http://manovich.net/index.php/projects/trending-the-promises-and-the-challenges-of-big-social-data>. Acesso em: 19 set. 2019
- MAYER-SCHONBERGER, V.; CUKIER, K. Big Data: como extrair volume, variedade, velocidade e valor da avalanche cotidiana. Rio de Janeiro: Campus, 2013.
- MILOSAVLJEVIĆ, M. VOBIČ, I. Our task is to demystify fears: analysing newsroom management of automation in journalism. *Journalism*, [s. l.], p. 1-19, 2019.
- OSTERWALDER, A.; PIGNEUR, Y. Business Model Generation Inovação em Modelos de Negócios: um manual para visionários, inovadores e revolucionários. Rio de Janeiro: Alta Books, 2011.
- REITER, E.; DALE, R. Building Natural Language Generation Systems. Cambridge: Cambridge University Press, 2000.
- RUÃO, T. Organização Comunicativa: teoria e prática em comunicação organizacional. Braga: CECS, 2016.
- SARAIVA, E. V.; CARRIERI, A. de P.; AGUIAR, A. R. C.; BRITO, V. da G. P. Um "pas de deux" da estratégia com a arte: as práticas do Grupo Corpo de balé. *RAC*, [s. l.], v. 15, n. 6, 2011.
- SOMMERVILLE, I. Engenharia de Software. Tradução Luiz Cláudio Queiroz. 10. ed. São Paulo: Person Education do Brasil, 2018.
- SOMMERVILLE, I. Engenharia de Software. Tradução: Selma Shin Shimizu Melnikoff, Reginaldo Arakaki, Edilson de Andrade Barbosa. 8. ed. São Paulo: Pearson Addison Wesley, 2007.
- STRAUSS, N. Financial journalism in today's high-frequency news and information era. *Journalism*, [s. l.], p. 1-18, 2018.
- TABARY, C.; PROVOST, A.-M.; TROTTIER, A. Data journalism's actors, practices and skills: a case study from Quebec. *Journalism*, [s. l.], p. 1-19, 2015.
- VOLK, Sophia Charlotte; BERGER, Karen; ZERFASS, Ansgar; BISSWANGER, Luisa; FETZER, Marcus; KOHLER, Karolin. How to play the game: Strategic tools for managing corporate communications and creating value for your organization. Leipzig, Germany: Academic Society for Management & Communication, 2017.
- WÖLKER, A.; POWELL, T. E. Algorithms in the newsroom? News readers' perceived credibility and selection of automated journalism. *Journalism*, [s. l.], p. 1-18, 2018.
- WOLTON, D. Informar não é comunicar. Porto Alegre: Sulina, 2011.
- ZERFASS, A.; VERCIC D.; NOTHHAFT, H.; WERDER, K. P. Strategic Communication: Defining the Field and its Contribution to Research and Practice. *International Journal of Strategic Communication*, [s. l.], v. 12, n. 4, p. 487-505, 2018. DOI: 10.1080/1553118X.2018.1493485.

RESUMO:

Esta pesquisa objetiva desenvolver o *Minimum Viable Product* (MVP) de uma plataforma em nuvem que transforma dados estruturados em textos automatizados, para potencializar a produção de conteúdo nas Corretoras de Valores Mobiliários. Trata-se de Pesquisa Aplicada que combina diferentes técnicas para o desenvolvimento do MVP. Como percurso teórico para embasar o desenvolvimento do produto, perpassa a comunicação estratégica nas corporações, assim como a adoção de algoritmos e a automação de conteúdo na produção jornalística. Como alguns dos resultados alcançados, temos a lógica de infraestrutura macro do funcionamento do *newsbot* e os principais elementos para as inúmeras camadas de desenvolvimento de que uma aplicação dessa natureza necessita.

PALAVRAS-CHAVE: Comunicação estratégica; Estratégia organizacional; Conteúdo automatizado; *Newsbot*; Algoritmos.

ducción de contenido de Corretajas de Valores. Se trata de una Investigación Aplicada que combina diferentes técnicas para el desarrollo de MVP. Como camino teórico para apoyar el desarrollo del producto, traspasa la comunicación estratégica en las corporaciones, así como la adopción de algoritmos y automatización de contenidos en la producción periodística. Como algunos de los resultados alcanzados, el trabajo presenta la lógica de infraestructura macro de la operación del *newsbot* y los elementos principales para las innumerables capas del desarrollo que una aplicación de esta naturaleza necesita.

PALABRAS-CLAVES: Comunicación estratégica; Estrategia organizacional; Contenido automatizado; *Newsbot*; Algoritmos

RESUMEN:

Esta investigación objetiva desarrollar el *Minimum Viable Product* (MVP) de una plataforma en la nube que transforma datos estructurados en textos automatizados, para potenciar la pro

