SOURCES, MANAGEMENT AND SCIENTIFIC DISSEMINATION: report on the development of a search tool for the Press

FONTES, GESTÃO E DIVULGAÇÃO CIENTÍFICA: relato do desenvolvimento de uma ferramenta de busca para Imprensa

FUENTES, GESTIÓN Y DIFUSIÓN CIENTÍFICA: informe sobre el desarrollo de una herramienta de búsqueda para Prensa

Rodrigo Eduardo Botelho-Francisco
PhD in Communication Sciences with habilitation degree (Livre-docência title in Brazil) in Information and Technology at the University of São Paulo (USP). Professor at the Department of Science and Information Management at the Federal University of Paraná (UFPR). Member of the UFPR Agência Escola. rodrigobotelho@ufpr.br.

0000-0002-4772-9398

Fabia Cristiane Ioscote
Doctoral Student and Master in Communication at UFPR. Journalist under graduated from Universidade Tuiuti do Paraná (UTP). Scholarship student Capes; Scholarship Student at UFPR Agência Escola (2018-2020). fabia.ioscote@ufpr.br.

0000-0002-6318-026X

Antonio Rodrigues Barros
Master in Science, Management and Information Technology and Technologist in Information Technology by UFPR. Information Technology Technical Public Servant of the permanent staff of UFPR. Technical scholarship at the UFPR Agência Escola. antonio@ufpr.br.

0000-0003-1929-1402

Matheus Dias Galdino Soares
Student at the UFPR's Communication Master's Program. Journalist undergraduated by the State University of Ponta Grossa (UEPG). Scholarship Student at UFPR Agência Escola (2020-2021). aomatheusao@gmail.com.

0000-0002-3078-3005

ABSTRACT:
This article aims to present a report of the development of the *Guia de Fontes @UFPR*, a search tool for scientific information created by the Agency School of Public Communication and Scientific and Cultural Dissemination (AE) of the Federal University of Paraná (UFPR). In this context, the paper discusses information and scientific content management from an interdisciplinary point of view and describes the development of the software covering Journalism, Information Management and Computing, areas of knowledge in which the methodological principles of the *Guia de Fontes @UFPR* are based. In conclusion, the work also presents the results under the Scientific Popularization.

KEYWORDS: Information sources; Content Management; Web development; Journalism; Scientific Popularization.

Introduction

This article aims to present a report on the development of the *Guia de Fontes @UFPR*[1], a search tool for scientific information sources developed by the *Agência Escola de Comunicação Pública e Divulgação Científica e Cultural* (School of Public Communication and Scientific and Cultural Dissemination Agency - AE) of the Federal

---

[1] Available at: www.guiadefontes.ufpr.br.

University of Paraná (UFPR). Specifically, it aims to discuss scientific information from an interdisciplinary point of view; discuss content management of this nature; and describe the actual software development itself.

A search tool is understood, in this article, as an instrument for locating informational resources (CENDÓN, 2001), designed to provide results based on keywords defined by an interactor in documents or in a specific database or even on the Internet. Also known as search engines or directories are popularly recognized in services provided by platforms such as Google and Yahoo.

From the point of view of the Press, a Source Guide is a search engine used in the process of locating experts who can collaborate in the production and writing process of news or reports, providing interviews or information that allow the journalist to understand concepts or facts, besides confronting opinions. In this sense, the source is considered an authority on the topic being consulted.

In journalism, sources are a matter of fundamental seriousness, the cultivation of which costs media and communication professionals time and money. Pinto (2009) reports that they have their own databases with the most accessible sources. For Traquina (2005), there is a news network in which human sources are inserted and distributed. Therefore, a guide that brings together these agents in a more logical and easily accessible way becomes a relevant ally of efficiency in professional routines, enhancing the reach of the news network.

The construction of Source Guides, however, faces problems such as data availability and reliability, updating and costs. This happens because it depends on constantly gathering information about people and their knowledge, which requires the insertion and manual updating of data in a structured base or its obtaining from other sources in an interoperable way. The fact is that information holders are not always available to participate in this registration and update process at a reasonably adequate pace. On the other hand, involving third parties in the process is costly and requires great care to ensure the quality of the registration.

Faced with this problem, *Guia de Fontes @UFPR* was developed by the *Agência Escola* from a perspective of using open scientific data, as it is based on the interoperable use of data from the Currículo Lattes, "a national standard for recording past and current life of the country's students and researchers, and is now adopted by most of the country's development institutions, universities and research institutes" (CNPq, [s.d.]).

AE, maintainer of the *Guia de Fontes @UFPR*, is a technical-scientific project, involving the Arts, Communication and Design Sector (*Setor de Artes, Comunicação e Design-SACOD*), the Communication and Marketing Superintendence (*Superintendência de Comunicação e Marketing - SUCOM*) and the Foundation of the Federal University of Paraná (*Fundação da Universidade Federal do Paraná - Funpar*), with the objective of carrying out a work of positioning the University as a relevant actor in the regional, national and international scenario by giving greater visibility to the scientific and cultural production developed therein or stimulated by it.

In this perspective, the proposal of the *Guia de Fontes @UFPR* is justified in view of the AE's Public Communication approach, which maintains as a principle the valuation of information of public interest, which interferes with social development. In addition, it is based on the perspective of technological innovation, one of AE's three axes of action, aimed at the development and improvement of digital solutions for Information Management, facilitating access by internal and external audiences to information about UFPR.

From a methodological point of view, the development of the *Guia de Fontes @UFPR* is centered on an interdisciplinary vision, involving knowledge of Journalism, Content Management and Computing. In this article, to present the report, the discussion is based on the perspective of information sources for these different areas of knowledge, on the different aspects involved in the representation and retrieval of information and on selected aspects of the software development process.

Information sources

Information sources vary according to their uses, intentions, characteristics and formats. *Guia de Fontes @UFPR* seeks, in this sense, to use sources according to their definitions by Information Sciences (IS), Journalism and by Computer Science.

Although *Guia de Fontes @UFPR* has as its main function to allow journalists and reporters a quick contact with specialized researchers from different areas of knowledge, the platform is guided by the idea of Information Sources, as it gathers different data from different sources and supports to offer a solution to the challenges of media work.

In IS, the sources of information are people, institutions and publications that can be classified from the perspective of location into primary, secondary and tertiary (ALBRECHT & OHIRA, 2000). The primaries are the original information, coming from the texts of articles, theses, dissertations, books and other publications. The secondary

would be the resources that gather and organize the primary information, such as databases, indexes and directories. Tertiary companies, in turn, add value through the use of the two categories mentioned above, facilitating the location and access to such information. "They represent the starting point for collection actions" (ALBRECHT & OHIRA, 2000, p. 140).

There are several areas that use source guides for their research activities, such as History, Library Science, among other specificities of knowledge. Archivology, for example, uses public databases or guides as proprietary sources to carry out its research. An example is the historical source guide developed by Silva (2011), which focused on the prison system in Rio de Janeiro between 1830 and 1890.

Source Guides are not always interactive, with a software interface. There are several examples in which the guides are closed files in printed or e-book formats, such as the Guide of Primary Sources of the Observatory of Museums and Cultural Centers of Fiocruz (PEREIRA & KÖPTCKE, 2008).

Journalists, on the other hand, use different sources in their professional activities. The selection of the origin, nature and characteristics of the sources will be determined by the type of agenda to be investigated, the support in which the news will be published, the journalistic genre in relation to which the material will be prepared and other factors that may be external to these criteria. Pinto (2009) defines the sources according to their nature: they can be human, documentary or statistical. With regard to human sources, the author defines four main types. They are: technicians and specialists; informants; news characters and analysts not directly involved with the fact. Traquina (2005), in his turn, states that human sources are classified by journalists based on their interests when commenting on an agenda. These interests are associated with authority, productivity or credibility. These would, in theory, be the counterpart for a source to participate in the development of journalistic content.

A Source Guide, in this sense, provides journalists and reporters with the contact of sources who serve mainly as experts, since they have specific knowledge about a certain subject and provide data in an identified way, or even as external analysts. The latter participate as "in off" sources (PINTO, 2009), being able to guide the reporter in the production of an agenda, but without necessarily appearing identified as a source that provided data or interviews.

Pinto (2009) suggests that journalists build a database of the most recurrent sources of their work, in order to facilitate an eventual contact, which must be made with the agility that the guidelines may require.

When it comes to the approximation between scientific production and Journalism for the broad public, Nascimento and Sommer (2005) point out that there are difficulties for reporters in translating information from scientific research to a readership that is not familiar with these topics. Much of this is due to the difficulty in accessing scientific sources of information that contribute to the daily work of investigation and reporting. Therefore, a source guide acts significantly to create an interface that allows journalists quick, systematized and simplified access to the most current scientific sources that are best suited to deal with different topics.

In the context of Computing, data sources represent repositories that can be accessed, grouped, processed and indexed in order to generate relevant information for the interactors who use a given system. Thus, it appears that data sources provide storage, as in the case of databases, which, according to Elmasri and Navathe (2005, p. 10) are "a collection of related data. Data are facts that can be recorded and that have an implicit meaning". The authors exemplify that these data can be names, telephone numbers and addresses. However, they are aware that this definition can be too general. In other words, a set of words used in a text that are related to each other could mistakenly assume a database concept. Therefore, Elmasri and Navathe (2005) establish some implicit properties that make the use of the term database more restricted for computing. They are: 1) representation of some aspects of the real world (mini world)[2]; 2) logical and coherent collection of data with some pertinent meaning; 3) designed, built and populated by data that meet a specific proposal.

Elmasri and Navathe (2005) state that data organized randomly, and that do not have defined user groups and interests, cannot be conceived as a database. In summary, "a database has some sources from which data are derived, some levels of interaction with real world events and an audience effectively interested in its contents" (ELMASRI & NAVATHE, 2005, p. 10).

Albrecht and Ohira (2000) point out that, at times, the literature presents confusion related to terminologies. Thus, it is common to find the terms database and database as synonyms. As a definition, a database gathers a set of databases, while a database is made up of data that are interrelated, in an organized way, in order to allow the retrieval of information. Therefore, "the purpose of a database is to provide

---

[2] Mini world or Universe of Discourse (UoD) refers to a structured collection of related data.

up-to-date information (structural resources), accurate and reliable (not giving half-information) and according to demand (offering what the user needs)" (ALBRECHT & OHIRA, 2000, p. 33). Therefore, an efficient database needs to store and at the same time provide mechanisms for an interactor to be able to find exactly what one is looking for.

In the academic field, in order to study and understand the various phenomena and knowledge in the various spheres of knowledge, Science also uses sources using procedures that are somewhat similar to journalistic research, but with methodological rigor and peer scrutiny. Proof of this is that many researchers use, for example, documental and bibliographic analysis and even interviews to gather information. While in Journalism, what the source says often becomes news, in Science there are other sieves and filters so that information can be part of scientific knowledge.

Sources for Science are in many places (SUBRAMANYAM & SUBRAMANYAM, 1981): technical and scientific communications; scientific societies; newspapers, magazines and periodicals; conference proceedings; dissertations, theses and research in process; patents; technical reports; standards and specifications; bibliographical sources; dictionaries; folders and yearbooks; tables and booklets; encyclopedias; translations; literature reviews; between others. But information can also be obtained from people, from interviews, focus groups and other controlled experiences with human beings. For Rodrigues and Blattmann (2014), the use of appropriate sources associated with assertive information management are essential for the generation of knowledge efficiently and with less risk. The scientific production itself has used systematized information sources in the form of source guides. Two examples are the Federal University of Minas Gerais (*Universidade Federal de Minas Gerais - UFMG*), which has the *Somos*[3] platform, which allows a search on its researchers, their skills, their research and their capacity within the institution. The *Acácia* platform[4], created by researchers from the Federal University of ABC (UFABC), also has the peculiarity of serving as a genealogy of academic production in Brazil, mapping researchers and their respective advisors and advisees, also making a calculation of fertility and fertility of the work that researcher so far, estimating how far the influence of that academic has gone.

Content Management

---

[3] http://somos.ufmg.br
[4] http://plataforma-acacia.org

One of the great challenges for working with data and information is the promotion of its integrated management. Isolated and without adequate representation and treatment, they are unable to represent anything. In this sense, Content Management arises, an area of studies that, from an organizational perspective, "supports organizations in the capture, organization and distribution of content originating from various sources and destined for different types of output devices" (PEREIRA & BAX, 2002, p. 03).

Given the contemporary context and aspects related to the web and the vast amount of content available daily in the digital environment, Content Management allows for the establishment of minimum requirements that offer a customized solution for organizations. Thus, Content Management is "a technological approach that arises from the explosion of multimedia content on the web and intranets and aims to allow the management of all stages, from content creation to publication" (PEREIRA & BAX, 2002, p. 10). From this perspective, with regard to content, it is characterized by texts, images, videos, audio files, graphics, among other documents such as spreadsheets, etc.

Considering that Content Management gains notability from the complexity of the web environment – in which the variety and quantity of available content become challenges for management –, in the same way, the content management systems (CMS, of the acronym, Content Management System) are important. According to Boiko (2005, p. 79, free translation), "a CMS can do much more than produce a website. It can cover any part of your content creation and organization system that you want"

Deploying a CMS in an organization requires flexibility and an easy-to-use system that facilitates content sharing. "When it comes to content, information and knowledge, each organization is unique and requires its own adaptation to its reality" (PEREIRA & BAX, 2002, p. 06). This implies that, in terms of content management, there is no single and universal solution that meets the specificities of organizations (PEREIRA & BAX, 2002). Furthermore, for development "it is preferable that a content management solution provides the 'bricks' *as atomic as possible for the required level of abstraction*" (PEREIRA & BAX, 2002, p. 06, *emphasis added*). In this way, whoever designs the system can act with flexibility in the assembly, reaching the most adequate solution for the organization's profile.

In this context of Content Management, many organizations have a simplistic view of the topic and possibly just look for answers or best practices to manage while continuing to carry out their work (BOIKO, 2005). For the author, the root of the

problem is the lack of a better understanding of the differences between data and content. Sometimes there is a conception that content is as simple as data.

The data, in this case, can be defined as a unit, a raw material that by itself does not add value. It is also understood as a basis for building information. About this concept, Boiko (2005) defines that "data consists of small fragments of computer information – numbers, words, images, sounds – that have much of the human meaning extracted from them" (BOIKO, 2005, p. 04, free translation). In this digital circumstance, the fact that we have a lot of data in this format does not imply that other records are not data (DAMA, 2017). In this way, "since we can capture so much information electronically today, we call many things 'data' that would not have been called 'data' in earlier times - things like names, addresses, dates of birth" (DAMA, 2017, p. 47, free translation).

The relationship between data and information has been widely discussed in the literature, placing these terms as inseparable, since they are vital in the daily routine of organizations. Therefore, data is called raw material of information and information is called data in context (DAMA, 2017). For authors such as Boiko (2005), information is conceived based on the human capacity to attribute meaning to data. Along the same lines, Choo (2003) defends more intrinsic characteristics of individuals in relation to information, and that this, as well as insights, arises in the mind. For the author, "the search and use of information is a dynamic and socially disordered process that unfolds in layers of cognitive, emotional and situational contingencies" (CHOO, 2003, p. 62). In short, data is what is obtained and information gains context from human subjectivity. From then on, according to Boiko (2005), content is defined by the sum of data and information.

With these differentiations made, the scientific content, therefore, will be characterized by the use of data of a scientific nature - these being numbers, words, images or sounds - allied to quantitative and/or qualitative information, with methodological rigor and a narrative that is part of the work carried out by researchers and scientists.

From a management perspective, information is a crucial element, especially in aiding knowledge and decision-making within organizations. Conceived in a cognitive instance, the information will be selected, identified and organized by the subject in a form of classification. The result of this process is called representation.

From a cultural, anthropological and also semiotic point of view, representation encompasses objects and concepts in the human mind. "The processing of these

representations before acting practically in the world, according to the result of this processing, is one of the most fundamental characteristics of humanity" (MARCONDES, 2001, p. 63). It is noteworthy that the process of searching for information by the individual occurs in a hierarchical manner, from a need for general information, to interactions with sources and information retrieval systems, to the final use of the information obtained (MARCONDES, 2001).

IS defines the representation associated with classification and information retrieval. This classification includes indexing, taxonomies, semantic trees and knowledge management. Research in this area discusses representation based on contemporary issues about the volume of information and the variety of devices available to access and search for information. Advanced Computation, in its turn, allows for a variety of proposals, languages and metadata for the purposes of cataloging, classifying and also indexing information. "Along with special HTML tags, META tags, metadata constitute a possible solution for the informational explosion brought by the Internet" (MARCONDES, 2001, p. 62).

In the organizational scenario, controlling and managing the volume of information is an ongoing challenge. Since information is understood as a valuable resource for decision-making, and is even essential to keep the organization competitive, the more you plan, the more information is needed (TEIXEIRA & VALENTIM, 2017). In this perspective, representation plays a fundamental role in the information economy. "In the information/knowledge economy, the representation must be so cognitively rich as to allow a user to infer the content of the document" (MARCONDES, 2001, p. 67). This is where representation is important, as it serves to save energy in the information search process. According to Marcondes (2001), when reading a document, for example, it is more economical to read the abstract than the full text.

Given the needs of users and organizations, significant changes in accessing and searching for information, such as the possibility of digital storage of large volumes of data, indexing languages, keyword research through search engines in text and voice, the use of URLs, different media and supports, connections and speed also challenge information retrieval systems. Agile and at the same time reliable systems are needed. "Thus, the information retrieval process needs efficient resources to provide relevant information" (TEIXEIRA & VALENTIM, 2017, p. 83-4). In this sense, information systems act as intermediaries between the needs for accessing and searching for information

and the representation of content. They are also considered linguistic communication tools.

But what if the data is represented and available in different databases and systems? It is in this context that the idea of interoperability arises, understood in the capacity of a system to communicate with another. For the Information Technology area, the concept is understood, in general terms, as the ability to exchange information between computers and programs from different manufacturers. Unlike integration, which refers to the connection process, interoperability allows two or more systems to work together in a communicational sense in exchanging data. The National Information Standards Organization (NISO) definition states that interoperability is the "ability of diverse systems, with different hardware and software platforms, data structures and interfaces, to exchange data with a minimum loss of content and functionality" (NISO, 2004, no p., free translation). According to Santarém Segundo et al. (2019), interoperability also means the ability to exchange data and information between different organizations regardless of the systems or platforms they use. From the point of view of Computing, "interoperability presents itself with many variables and requirements, ranging from the file structure to be exchanged to the functional models of computational access" (SANTARÉM SEGUNDO et al., 2019, p. 75).

In a scenario like the Internet, in which there is a high volume of information, complexity, structures and flow, interoperability is important. "It is useless for information to exist if those who need it do not know of its existence or if it cannot be found" (SAYÃO & MARCONDES, 2001, p. 26). In this context, the idea of transparency of Information Systems also emerges. In conceptual terms, "transparency is defined as the concealment from the user and application programmer of the separation of components in a distributed system" (COULOURIS et al., 2012, p. 23, free translation). It implies that the system must be perceived as a whole and that, from the user's point of view, it is not interesting to know about components, but rather, the final result that this system can offer.

In the area of Software Engineering, Cysneiros (2013) explains that there are numerous definitions about a transparent system, but the main point of transparency is the disclosure of information. "We believe that for the software to be transparent, the information it handles must be transparent (Transparency of Information)" (CYSNEIROS, 2013, p. 19, free translation). Thus, transparency and interoperability are essential for an information system to be efficient.

Software Development

Guia de Fontes @UFPR was developed with the aim of giving the press and society in general access to information about the specialties of Brazilian researchers. With the tool, in addition to facilitating journalistic work and promoting the production of quality content, it is intended to promote scientific dissemination and academic transparency.

The solution is being developed as a free software and from free or open source technologies, which means that there are dependencies that force its source code to be equally shared with the same logic. In doing so, *Agência Escola* prospects and develops a tool that it hopes can be shared with other higher education institutions, in the logic of contributing to public communication and scientific dissemination.

From a development standpoint, there are several aspects of Computer Science involved in creating and maintaining software. In this sense, Software Engineering encompasses "a process, a set of methods (practices) and a range of tools that enable professionals to develop very high quality software" (PRESSMAN & MAXIN, 2016, p. 14). However, for this article, for reasons of convenience, it was decided to discuss one of the main topics of project management, Requirements Analysis, in addition to the perspective of methodologies and methods and technologies adopted.

Guia de Fontes @UFPR project had an analysis of requirements based especially on the demands of the Communication Superintendence (SUCOM) and the AE of UFPR, which in the past years had similar initiatives, with the same aim of providing information about specialties and specialists at the University for the Press. However, in this analysis, it appears that the solutions were either focused on building local databases or on other "home-made" initiatives by the Press Officers team itself, focused on the use of popular data spreadsheet software, such as the Microsoft Excel.

From observation and dialogue with the teams, the functional requirements of the project were established, centered on three profiles of interactors, as described below:

- **Journalists:** search for keywords on research topics; search for researchers' names; visualization of data for contacting researchers (phone, email); summary view of researchers' curriculum; data update date visualization; access to the full lattes curriculum of researchers.
- **Communication advisors:** system login, with username and password; password recovery; user registration; list users; list researchers;

add/change researchers contacts; definition of contacts as public or private; managing (adding or removing) keywords from the researchers curriculum; sending a temporary password to researchers; synchronization and updating of researchers' data with the Lattes Curriculum.

- **Researchers:** login to the system, with username and password; password recovery; add/change your contacts; definition of contacts as public or private; managing (add or remove) keywords from your resume; synchronization and updating of data from the Lattes Curriculum.

The non-functional requirements, in turn, are related, in the *Guia de Fontes @UFPR* project, to interoperability with Institutional Lattes; scheduled and on-demand data synchronization; database indexing to provide transaction speed and user response; usability, unnecessary training, tutorial or manual for system operation; accessibility; use of free and/or open source technologies in the development process; and open source for other Brazilian HEIs.

The development itself was centered on the agile philosophy, which in computing is understood as "customer satisfaction and early incremental delivery; small and highly motivated project teams; informal methods; minimal software engineering artifacts; and, above all, , simplicity in general development" (PRESSMAN & MAXIN, 2016, p. 66). Furthermore, from a principled standpoint, "they prioritize delivery over analysis and design (although these activities are not discouraged); they also prioritize active and continuous communication between developers and customers'' (PRESSMAN & MAXIN, 2016, p. 66).

In this process, favoring the acceleration of communication and collaboration among all participants, Scrum was adopted, an agile software development method used "to guide development activities within a process that incorporates the following methodological activities: requirements, analysis, design, evolution and delivery" (PRESSMAN & MAXIN, 2016, p. 78). In this sense, it is important to highlight three sets of development activities: backlog (priorities of requirements and project features), Sprints (work unit, measured in the time adjusted to meet the goals established in the project), Scrum meetings (short meetings, with updates the progress of the work of each team member) and Demos (incremental delivery of implemented features for customer evaluation).

Also highlighted are some aspects of the Information Architecture of the *Guia de Fontes @UFPR*, which are important to understand how the interface was designed.

This architecture is understood from the knowledge directed to the treatment of "informational, structural, navigational, functional and visual aspects of digital information environments through a set of methodological procedures in order to assist in the development and increase of usability" (CAMARGO & VIDOTTI, 2011, p. 24).

In terms of system organization and labeling, it is important to note the search engine, which occupies a prominent place in the Guide's interface, given the nature of the software and its centrality for navigating the system; as well as the organization of secondary, user-centric or content-centric menu items.

Focused on the user, the Guide was developed with a top menu that provides access to informative content for each user profile (Journalist, Communication Advisor and Researcher). In addition, the login also allows for a view with specific functionalities for both advisors and researchers, ensuring that everyone can have quick and easy access to information directly related to what the Guide provides for each profile.

Focused on content, a menu was developed, also at the top of the site, with labels based on the names of the major areas of Science (Exact and Earth Sciences; Biological; Engineering; Health; Agrarian; Applied Social; Human; Linguistics, Letters and Arts) and that function as editorials, giving access to the latest articles published by the Institution's researchers. In this case, the CAPES Knowledge Areas Table was used, a corporate taxonomy that "presents a hierarchy in four levels, from the most general to the most specific, covering nine major areas in which the 48 assessment areas are distributed" (CAPES, 2017).

Figure 1 *Guia de Fontes @UFPR's* screen where the composition of the brand and the menu structure can be seen.
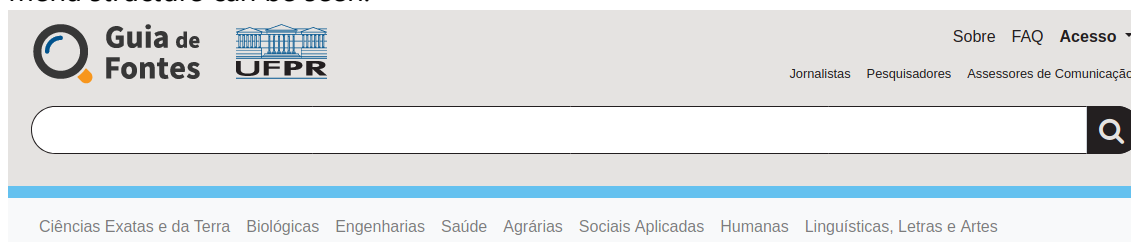
Source: the authors (2021).

Also in relation to the interface, the development and layout of the Guia's logo stands out, which, created by the Institutional Communication and Marketing team at Sucom UFPR, was based on a mixed brand, combining an image with a nominal part (in this case the name of the software) and a corporate brand (in this case, UFPR). The

strategy reflects a non-functional requirement of the software, since when the intention is to share it as free code, the visual disposition of the brand was designed to respect the aggregation of the brand in each organization that will be used. Thus, you get the *Guia de Fontes @UFPR*, just as you can get any other *Guia de Fontes* @anotherorganization.

Figure 2 Interface with the internal view, after login, of a researcher in the *Guia de Fontes @UFPR*.

Source: the authors (2021).

The Guide's data come from institutional databases of UFPR and institutional Lattes, a set of data about the University from the CNPq Curriculum Lattes, specifically areas of knowledge, research lines, text informed by the researcher and frequency of a set of words associated with the scientific and technical production registered in the Curriculum. As they are indirectly related to the researcher's expertise, keywords from judging panels, editing, event organization, guidance, participation in events and translations are excluded from the search.

Only data from active and retired faculty staff are available in the Source Guide. About them, as well as the Lattes Platform, the Guide only makes public information available, informed to the platform by the researcher himself when he fills out his curriculum. In this sense, they are his full responsibility.

This integration is one of the Guide's main differentials, as it ensures interoperability for data reuse, automation and continuous updating, as well as transparency in the Database construction process. This prevents both researchers and communication advisors from having to manually enter data in the *Guia de Fontes*, replicating an intellectual effort that already exists in updating the Lattes Curriculum, a nationally recognized scientific database that is widely used in the Brazilian academic scenario.

In terms of programming, the *Guia de Fontes @UFPR* was developed with Java and MySQL, both in version eight. The first is object-oriented programming language created in the 90s by Sun Microsystems, since 2009 is maintained by Oracle Corporation. The second is a Database Management System (DBMS) created in the 90's, widely used on the Internet as a data repository for websites and information systems, since 2009 it is also maintained by Oracle Corporation.

MySQL was chosen as the main database because it is a DBMS widely used on the web and is free to use, that is, there is no need for license costs. In addition, it is a database widely used at UFPR.

The use of MySQL as a single database to provide the services of the Source Guide, however, involved the integration of heterogeneous databases facilitated by the interoperability between systems, as can be seen in Figure 3. The main base of the Guide, called the metabase, it was built from the following data sources:

- **Oracle Database Management System (institutional data source of UFPR)** - used to obtain the minimum data (name, CPF and e-mail) of the Institution's professors;
- **Lattes PRPPG110**[5] - remote service that allows obtaining the researcher's identifier in the CNPq, using the CPF as a parameter;
- **Lattes PRPPG2** - remote service that allows obtaining the Lattes curriculum in XML[6] format, using the researcher's identifier as a parameter;

---

[5] PRPPG is the acronym for Pro-Rectory of Research and Graduate Studies at UFPR, which in this case is the keeper of Lattes Institucional's services.

- **Pages of Scientific Journals and Journals (where the articles are published)** – search for meta information on the pages (Dublin Core[7] and Open Graph[8]) in order to obtain the summary of each article, using the DOI[9] and Home Page fields of each publication to locate the journal pages.

Figure 3 Architecture of the *Guia de Fontes @UFPR*.
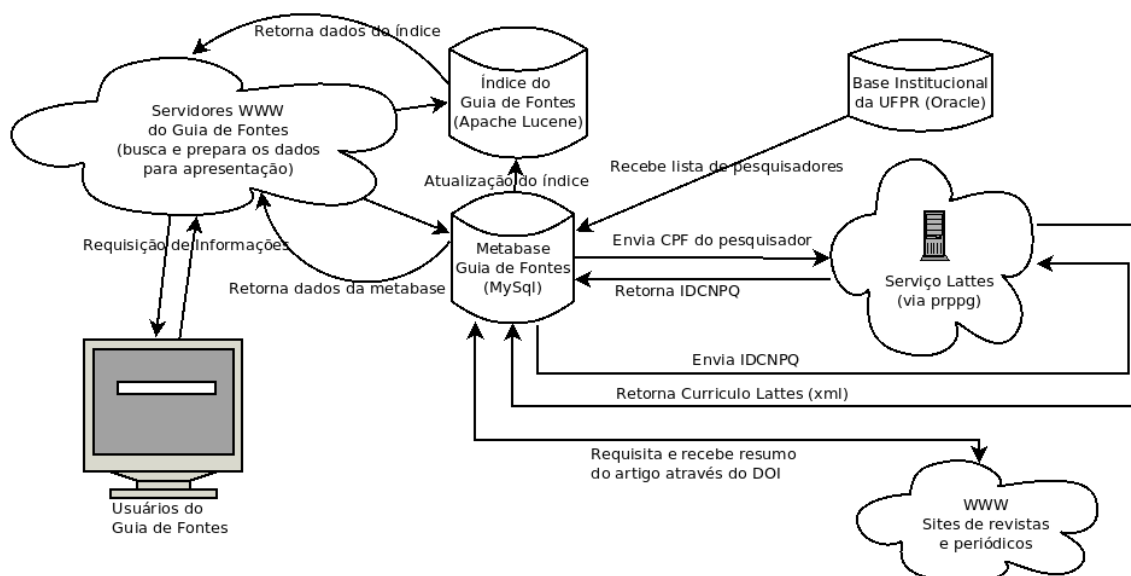
Source: the authors (2021).

Metabase, in the project, can be understood as the result of the integration of several data sources condensed and conveniently arranged in a single base. This, in turn, can be indexed and optimized to allow the extraction of relevant information for the purpose for which it is intended.

Integration takes place asynchronously, being periodically updated at eight-day intervals for resume data and on demand for publication summaries. This means that when publications in the *Guia de Fontes* are accessed, the abstracts are searched asynchronously and stored for future searches.

---

[6] XML stands for Extensible Markup Language, a markup language recommended by the World Wide Web Consortium (W3C) for describing different types of data, with the main objective of facilitating the sharing of information on the Internet.

[7] Dublin Core is a set of standardized metadata used to describe resources on the web, such as videos, images, texts, etc

[8] Open Graph is a protocol that defines a set of metadata to facilitate the sharing of resources (videos, images, texts etc) on social media such as Facebook, Instagram, Twitter etc.

[9] Digital Object Identifier (DOI) is a unique identifier registered at http://doi.org.br. The service, a pointer that allows each document (book, article or other type of publication) to be unique in the digital medium, is maintained by the International DOI Foundation (IDF).

Once built, the metabase is submitted to an indexing routine based on the following criteria: name of the researcher; resume summary; agglutinated keywords – all keywords used in the researcher's curriculum; research lines; and area of expertise; in addition to publication data, such as title, keywords, areas of knowledge, abstract, etc. This data is indexed by Apache Lucene[10] and is available in an index repository managed by Apache Lucene's own search library.

In this case, Apache Lucene's function is to facilitate relational database indexing in its natural form, i.e. tables, columns, relationships, etc.

As a web system, the *Guia de Fontes* is made available on the web through the Apache HTTP and Apache Tomcat server pages, solutions developed by the Apache Software Foundation and distributed as free software. These work together to enable the exposure of the Source Guide on the web and are made available and maintained in UFPR's service infrastructure.

Other dependencies of the development process are centered on technologies and frameworks such as the Model View Controller (MVC) Springboot in its version two, as well as several other technologies that are part of the MVC framework; and Hibernate Search[11], which has great prominence in the project as a search engine used.

Gitlab[12] is the version manager used to manage and store the source code of *Guia de Fontes*, which is hosted on the UFPR infrastructure. Its distribution is expected to take place under the GNU General Public License (GPLv3)[13], which among other things preaches freedom to use, modify and share source code.

Guia de Fontes and Scientific Dissemination

Since *Guia de Fontes* is conceived as a source of scientific information, serving as a concrete solution for the Press and for society in the search for information of this nature, the democratic bias of this action stands out. Thus, the proposal raises debate and contributes to Scientific Dissemination (SD), which, in turn, has the fundamental role of "democratizing access to scientific knowledge and establishing conditions for the so-called scientific literacy" (BUENO, 2010, p. 1).

About SD, Bueno (2010) draws attention to the public profile related to the process of circulating scientific information. In Scientific Dissemination, this audience "is primarily uninitiated, that is, they do not necessarily have technical-scientific training

---

[10] https://lucene.apache.org.
[11] http://hibernate.org/search.
[12] https://gitlab.ufpr.br.
[13] https://www.gnu.org/licenses/quick-guide-gplv3.html.

that allows them, without greater effort, to decode technical jargon" (BUENO, 2010, p. 2). Thus, DC conveys information in a way that it is accessible, in terms of language, to the lay public.

Figure 4 Presentation of result for keyword search in the *Guia de Fontes @UFPR*.

Source: the authors (2021).

As a counterpoint, from the perspective of the public in relation to the perception of Science and Technology in Brazil, a survey carried out by Delabio et al. (2021) reveals that most respondents believe that Science and Technology bring more benefits than harm to humanity. However, the authors highlight contrasts evidenced in the study, such as the lack of public interest in the topic and participation in decisions related to the direction of Science, in addition to a belief that some people are incapable of learning Science. Furthermore, even believing in benefits, they do not know which ones (DELABIO et al., 2021). In this way, the importance of Scientific Dissemination is reinforced by a social role also as Public Communication that has a citizen perspective and includes themes of collective interest (DUARTE, 2007).

Already seeking to understand the views of Brazilian scientists, a study conducted by Massarani and Peters (2016) mapped how they observe the media and their relationship with journalists. In general terms, the results of the work showed that there is an agreement among scientists that the communication of their research needs to be

19

adapted in view of the specificities of Science and technical jargon. "The answers suggest that scientists accept, at some level, that they must adapt their way of speaking to journalists, but only to a certain extent" (MASSARANI & PETERS, 2016, p. 1170-1, free translation). Thus, scientists believe that a more relaxed language should be left to reporters, understanding that this can be a strategy for the Press in maintaining the public's attention.

Another important point regarding the study carried out by Massarani and Peters (2016) concerns the availability of scientific publications to journalists. According to the authors, the interviewed scientists reported that in the last decade they had easy access to articles in international journals through the *Capes Portal*. "But this is not the case for journalists" (MASSARANI & PETERS, 2016, p. 1172, free translation). In this sense, it is noteworthy that although there are some initiatives in the Brazilian context, such as Scielo[14], the delay in making new issues available by periodicals is "a considerable obstacle for journalists who need up-to-date information due to the nature of journalistic production " (MASSARANI & PETERS, 2016, p. 1172, free translation).

In addition to the difficulties regarding Scientific Dissemination such as the public's perspective, the relationship between scientists and journalists and the availability of scientific publications, journalistic routines such as investigation, contact with the source, checking, revision and editing flows must also be considered. Added to this, the independent voice, a result after all the processes of journalistic work. Also as pointed out by Massarani and Peters (2016), there is an understanding on the part of scientists that journalists should consult them before publishing a story in order to avoid mistakes. However, as highlighted by the authors, this is an idea strongly rejected by journalists. In this sense, journalists understand that their role is to mediate information between things in the world and society. Given the professional journalistic culture, it would not be pertinent to review a story by the interviewee/specialist/scientist due to the technical specificities of the subject or theme. However, Bueno (2010), in turn, points to a possible spectacle of the news, especially when the journalist is not able to decode the specialized discourse.

These, as well as all aspects of access and search for scientific information together with communication about Science – important for the constitution of Scientific Dissemination – make a Source Guide relevant as it is a tool that can bring scientists together, journalists and society in general. Above all, as a Source of Information, a guide of this nature can be assumed in view of the concept of

---

[14] http://scielo.br.

Knowledge Resources (KR) (CASSOTTA et al., 2017), which are defined as means or tools intended to store, share and promote the reuse of scientific knowledge. In this sense, the main purpose of a KR "is the integration and/or relationship between authors, researchers, professors and students to deal with information of common interest" (CASSOTTA et al., 2017, p.19).

The lack of agile mechanisms that facilitate the organized search and the dialogue between journalists and scientific sources compromise the efforts of Public Communication and Scientific Dissemination. Faced with this challenge, the *Guia de Fontes @UFPR* favors a rapprochement between scientists and journalists, being a means or a tool capable of contributing to journalistic routines - from the investigation and contact with sources - while contributing to access and search for scientific content, whether by specialists or society in general.

Final Considerations

This article aims to present a report on the development of *Guia de Fontes @UFPR* and also to discuss the perspective of interdisciplinarity between the areas that provide the necessary knowledge for discussion on the search for scientific information from an applied research perspective.

As a Source of Information, the Guide can be understood from different perspectives, both as a Database, and from the idea of a source of information, as well as discussed by IS and by Journalism itself. From a management point of view, it is important to observe the complexity of the scientific content, as well as the aspects of representation and retrieval of information of this nature in an organizational context dependent on interoperable technologies. In addition, the idea of transparency is highlighted, since the interactor, when using search engines to find information of this nature, expects to interact with the system without the need for specialized knowledge for its operation.

Another highlight about the representation of information is that it assumes intrinsic characteristics on the part of individuals. When accessing a system to perform a search, there is an inference about the results that determine what matters and what doesn't. In this sense, the results interface of the *Guia de Fontes @UFPR* has an Information Architecture based on usability, which consequently reflects on the information economy (MARCONDES, 2001).

As a result, *Guia de Fontes @UFPR* facilitates the production routine of journalists and press officers with regard to contact with human sources, in this case, UFPR

researchers. The search for keywords meets the demand with the necessary agility in a contemporary context in relation to the Internet. Above all, the way in which the Guide was designed helps both the work of the Press and society in general, since it can be used in the search for specific themes or subjects that are at the same time of collective interest. In addition, given the volume and variety of information in different periodicals that can sometimes be an obstacle in obtaining or surveying publications, a Source Guide is a means that can bring scientists closer to journalists.

Naturally, the account of the development of the Guide in this article is limited to the context and limitations of the format itself. It is known that discussions on scientific dissemination in contemporary times require an in-depth, reflective and critical debate, based both on a perspective of Public Communication and Democratization of Science, as well as on a conception of Science and Technique that allows opening the "black box" of technology, and present both the principles of its design as well as the controversies involved in the process. Treading this path, however, requires a posture of applied research, which, it is believed, is visible in the reported development and prototyping process.

The idea of functional prototypes, in its turn, is important for applied research because it allows working with experiments, through the gradual release of small portions of the system to be tested, putting concepts to the test, revisiting and revising them constantly, with a more effective participation of the actors involved. In this sense, the *Guia de Fontes @UFPR*, based on the development methodology adopted, has, at the end of this article, a versioning (version 3.0.0) and a backlog that prove an agile development, at the same time as a scenario of future works, including new functionalities for integration with scientific information systems, more functionalities for journalistic work, mailing management, notifications, among others.

References

ALBRECHT, R. F., & OHIRA, M. L. B. (2000). Bases de dados: metodologia para seleção e coleta de documentos. Revista ACB, 5(5), 131-144.

BOIKO, B. (2005). Content management bible. Indianapolis: Wiley Publ.

BUENO, W. C.. (2010). Comunicação científica e divulgação científica: aproximações e rupturas conceituais. Informação & Informação, 15(1esp), 1-12.

CAMARGO, L. S. de A. de; & VIDOTTI, S. A. B. G.. (2011). Arquitetura da informação: uma abordagem prática para o tratamento de conteúdo e interface em ambientes informacionais digitais. Rio de Janeiro: LTC.

CAPES. ([s.d.]). Tabela de Áreas de Conhecimento/Avaliação. Brasília, CAPES. Recuperado de: <https://bit.ly/2Y4SCeF>.

CASSOTTA, M. L. J.; LUCAS, A.; BLATTMANN, U.; & GODOY VIERA, A. F.. (2017). Recursos do conhecimento: colaboração, participação e compartilhamento de informação científica e acadêmica. Informação & Sociedade, 27(1).

CENDON, B. V.. (2001). Ferramentas de busca na Web. Ciência da Informação, 30, 39-49.

CHOO, C. W.. (2003). A organização do conhecimento: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Editora SENAC.

CYSNEIRO, L. M.. (2013). Using i* to Elicit and Model Transparency in the Presence of Other Non-Functional Requirements: A Position Paper. In iStar (pp. 19-24).

CNPQ. ([s.d.]). Sobre a plataforma Lattes. Recuperado de: <http://lattes.cnpq.br>.

COULOURIS, G.; DOLLIMORE, J.; KINDBERG, T.; & GORDON, B.. Distributed Systems. Concepts and design. 5. ed. Boston: Addison-Wesley.

DAMA. (2017). Data Management Body of Knowledge DAMA-DMBOK®. 2. ed. Westfield: Technics Publications.

DELABIO, F.; PIAI CEDRAN, D.; MORI, L.; & MICHELLAN KIORANIS, N.. (2021). Divulgação científica e percepção pública de brasileiros (as) sobre ciência e tecnologia. Revista Insignare Scientia-RIS, 4(3), 273-290.

DUARTE, J.. (2007). Comunicação pública. (LOPES, Boanerges, Org.). Gestão em comunicação empresarial: teoria e técnica. Juiz de Fora: Multimeios.

ELMASRI, R.; & NAVATHE, S.. (2005). Sistemas de banco de dados. São Paulo: Pearson Addison Wesley.

MARCONDES, C. H.. (2001). Representação e economia da informação. Ciência da informação, 30(1), 61-70.

MARCONDES, C. H.; & SAYÃO, L. F.. (2001). Integração e interoperabilidade no acesso a recursos informacionais eletrônicos em C&T: a proposta da Biblioteca Digital Brasileira. Ciência da Informação, 30, 24-33.

MASSARANI, L.; & PETERS, H. P.. (2016). Scientists in the public sphere: interactions of scientists and journalists in Brazil. Anais da Academia Brasileira de Ciências, 88, 1165-1175.

NISO. (2004). Noções básicas sobre metadados. Recuperado de: <https://bit.ly/3ypwqrL>.

PEREIRA, M.; & KÖPTCKE, L.S.. (2009). Guia de Fontes Primárias: O Museu Nacional: seu público no Século XIX e início do Século XX. Rio de Janeiro: Fundação Oswaldo Cruz.

PINTO, A. E. S.. (2009). Jornalismo Diário: reflexões, recomendações, dicas e exercícios. São Paulo: Publifolha.

PRESSMAN, R. S.; & MAXIM, B. R.. (2016). Engenharia de software: uma abordagem profissional. 8. ed. Porto Alegre: AMGH.

RODRIGUES, C.; & BLATTMANN, U.. (2014). Gestão da informação e a importância do uso de fontes de informação para geração de conhecimento. Perspectivas em Ciência da Informação, 19, 4-29.

SANTAREM SEGUNDO, J. E.; SILVA, M. F.; & MARTINS, D. L.. (2019). Revisitando a interoperabilidade no contexto dos acervos digitais. Informação & Sociedade, 29(2).

SILVA, M. R. N.. (2011). O Arquivo e a Cibercultura. Guia de fonte sobre a história da prisão no Rio de Janeiro: 1830-1890 (pp. 1-12).

NASCIMENTO, M. de J.; & SOMMER, C.. (2005). Jornalistas e o acesso à informação. Em Questão, 11(2), 425-452.

SUBRAMANYAM, K.; & SUBRAMANYAM, B.. (1981). Scientific and Technical Information Resources. Boca Raton (EUA): CRC Press.

TEIXEIRA, T. M. C.; & VALENTIM, M. L. P.. (2017). Processo de busca e recuperação de informação em ambientes organizacionais: uma reflexão teórica sobre a subjetividade da informação. Perspectivas em Ciência da Informação, 22(4), 82-97.

TRAQUINA, N.. (2005). Teorias do Jornalismo, porque as notícias são como são. Florianópolis: Insular, 2. Ed..

RESUMO:
Este artigo tem por objetivo apresentar um relato do desenvolvimento do Guia de Fontes @UFPR, uma ferramenta para busca de informação científica criada no âmbito da Agência Escola de Comunicação Pública e Divulgação Científica e Cultural (AE) da Universidade Federal do Paraná (UFPR). Neste context, o trabalho discute a informação e a gestão de conteúdo científica sob um ponto de vista interdisciplinary e descreve o desenvolvimento do software contemplando o Jornalismo, a Gestão da Informação e a Compuação, areas do conhecimento nas quais estão pautados os princípios metodológicos do Guia de Fontes @UFPR. Por fim, o trabalho apresenta os resultados à luz da Divulgação Científica.

PALAVRAS-CHAVE: Fontes de Informação; Gestão de Conteúdo; Desenvolvimento web; Jornalismo; Divulgação Científica.

RESUMEN:
Este artículo tiene como objetivo presentar un informe de desarrollo de la *Guia de Fontes @UFPR*, una herramienta de búsqueda de información científica creada por la Agencia Escuela de Comunicación Pública y Difusión Científica y Cultural (AE) de la Universidad Federal de Paraná (UFPR). El trabajo hace una discusión de la información y de la gestión de contenido científica desde un punto de vista interdisciplinario. Además, describe el desarrollo del software a partir del Periodismo, de la Gestión de la Información y de la Informática, áreas del conocimiento en las cuales se buscan los principios metodológicos del *Guia de Fontes @UFPR*. En conclusión, el trabajo también presenta

24

los resultados bajo la discusión de Difusión Científica.

PALABRAS-CLAVES: Fuentes de información; Gestión de Contenidos; Desarrollo web; Periodismo; Difusión Científica