Article

Analysis of search engines of open access institutional repositories

Lucelia da Silva Almeida[1] iD, Thiago Magela Rodrigues Dias[2] iD

ABSTRACT

Introduction: Scientific communication in its new model of performance, openly enabling access to information, has contributed that society can resort to scientific production in an easier way, which has occurred due to its availability in open access journals and institutional repositories. In a scientific context, the retrieval of information provides access to intellectual production from the possibility of its search in retrieval systems. Objective: To analyze the information retrieval systems (IRS) based on the search tools of the Brazilian institutional repositories of theses and dissertations available in the BDTD and its databases, comparing them with the parameters established by IBICT. Methodology: Conducts bibliographic research on the theme and documentary research from access to the standards, guidelines, and manuals provided by IBICT. Results: It confirms that the institutions participating in the BDTD do not necessarily need to stick to the standards in a partial way established by IBICT, being necessary only that they assign to their repositories the mandatory metadata standards, along with the use of the OAI-PMH protocol and the use of XML technology so that the collection of data occurs for availability in the national BDTD portal. Conclusion: The information retrieval systems of the repositories have technological tools and specifications that confer navigation in the systems by the users, contributing to the best search experience and access to information relevant to their informational issues.

KEYWORDS

Information retrieval systems. Institutional repositories. Search engines. Open access.

| 1

Análise dos mecanismos de busca de repositórios institucionais de acesso aberto

RESUMO

Introdução: A comunicação científica em seu novo modelo de atuação, possibilitando o acesso à informação de forma aberta, tem contribuído para que a sociedade recorra as produções científicas de forma facilitada, o que se deu devido a sua disponibilização em periódicos e repositórios institucionais de acesso aberto. A recuperação da informação, em contexto científico, confere acesso a produção intelectual a partir da possibilidade de sua busca em sistemas de recuperação. Objetivo: analisar os sistemas de recuperação da informação (SRI) com base nas ferramentas de busca dos repositórios institucionais brasileiros de teses e dissertações disponibilizados na BDTD e suas bases de dados, comparando-os com os parâmetros estabelecidos pelo IBICT. Metodologia: Realiza pesquisa bibliográfica sobre o tema e pesquisa documental a partir do acesso aos padrões,

Authors' correspondence

[1] Universidade Federal de Santa Catarina, Florianópolis, Brazil / e-mail: lucellya.silva@gmail.com

[2] Centro Federal de Educação Tecnológica de Minas Gerais, Belo Horizonte, Brazil / e-mail: thiagomagela@gmail.com

diretrizes e manuais fornecidos pelo IBICT. Resultados: Confere que as instituições participantes da BDTD não necessariamente precisam ater-se aos padrões de forma parcial estabelecidos pelo IBICT, sendo necessário apenas que estas atribuam a seus repositórios os padrões de metadados obrigatórios, juntamente com o uso do protocolo OAI-PMH e o uso de tecnologia XML para que ocorra a coleta dos dados para disponibilização no portal nacional da BDTD. Conclusão: os sistemas de recuperação da informação dos repositórios possuem ferramentas tecnológicas e especificações que conferem a navegação nos sistemas por partes dos usuários, contribuindo para a melhor experiência de busca e acesso a informações pertinentes as questões informacionais dos mesmos.

PALAVRAS-CHAVE
Acesso aberto. Sistemas de recuperação da informação. Repositórios institucionais. Mecanismos de busca.

CRediT

| 2

JITA: IN. Open science.

Article submitted to the similarity system

1 INTRODUCTION

Scientific communication in its new model of performance, openly enabling access to information, has contributed that society can resort to scientific production in an easier way, which has occurred due to its availability in open access journals and institutional repositories.

Access to scientific productions enables their use and intertextuality in other productions, giving visibility and recognition to their authors. Access depends on retrieval systems that process queries made in their search engines, checking results from a list of related documents.

In a scientific context, the retrieval of information provides access to intellectual production from the possibility of its search in retrieval systems. The availability of a document free of charge in a database can be helpful for society. Still, without the possibility of access and concern with its relevance for a given user, it will not be liable for the solution of its informational problem.

Open access, in its greenway strategy (self-deposit in institutional repositories), is not only concerned with making scientific production available but also with access and use, that is, how to communicate it to society; this is related to sound management of the documentation produced in its environment (FERREIRA, 2009). In this way, the retrieval of information in a systematic way contributes to broad access to the information contained in an institutional repository (IR), enabling access to the documents contained therein.

Since information and knowledge are primary inputs in the "scientific and intellectual work process," free access by the researcher becomes a demand. Because the gains in quality and productivity increase as it becomes possible to select, filter, analyze, process, and recombine a more significant amount of information. And information and communication technologies have made this process increasingly powerful. (MACHADO, 2015). In this sense, the information retrieval process is a factor to be considered when promoting open access to information, bearing in mind the need to meet society's demand for information in general.

Based on this perspective, it is intended to work with the following issues: a) What are the search engines and tools implemented in the Brazilian institutional repositories of theses and dissertations made available by the Portal of the Brazilian Digital Library of Theses and Dissertations (BDTD), for the access to information? b) How do they relate to the parameters established by the Brazilian Institute of Information in Science and Technology (IBICT)?

This way, the general objective consists of analyzing the information retrieval systems (IRS) based on the search tools of the Brazilian institutional repositories of theses and dissertations available in the BDTD and its databases and comparing them with the parameters established by IBICT. And for specific purposes:

A) Talk about IRS and the tools that facilitate their access and dissemination in the context of open access institutional repositories;
b) Identify the search alternatives and other tools available for the databases of the institutional repositories participating in the BDTD;
c) Compare the systems' search tools with the parameters established by the IBICT.

This work is based on approaches that deal with IRS and information retrieval processes, specifically concerning retrieval in an open access system. For the structuring and consolidation of the research, we sought to analyze and describe the information provided by the pages about the BDTD, its creation process, its implementation, and the tools and other technologies it uses for the functional flow of the platform and the proposed objectives.

In this context, this article is structured into 5 sections. The first deals with this introduction. The second refers to the methodological procedures that led to this research. The third section and its subsections address discussions that deal with information retrieval, the systems that make it possible, and their relevance for accessing information in open access

| 3

repositories. They also highlight the open access movement, its purposes, origin, central concepts, and consolidation strategies.

The fourth section, together with its subsections, addresses the results of this research, describing the aspects corresponding to the observations and analyzes on the BDTD portal and the repositories of participating institutions. And as the last section, the final considerations regarding the development and result of this investigation.

2 METHODOLOGY

In the development of research, it is necessary to use methods and techniques that will guide the study and contribute to achieving the results of the defined research problem through the use of methodological procedures and tools that will be useful in the course of the study. "We can define a method as a way to reach a particular end. And scientific method as the set of intellectual and technical procedures adopted to achieve knowledge" (PRODANOV; FREITAS, 2013, p. 24). Therefore, bibliographical and documentary research was carried out regarding the procedures. As for the approach, we worked with qualitative research, and concerning the objectives, we developed a descriptive study.

Bibliographical research was necessary to contextualize the need for access to information and, consequently, the quality of its retrieval, contributing to the orientation and basis of the study. According to Marconi and Lakatos (2014), bibliographical research encompasses the survey of already published productions that address the subject addressed in the study, which can be found in the form of books, articles, theses, dissertations, etc., enabling the researcher to have knowledge about the approaches that are relevant to the theme of the study.

Documentary research, "[...] that carried out from documents, contemporary or retrospective, considered scientifically authentic [...]" (GERHARDT *et al.*, 2009, p. 69), can be confused with bibliographic research due to its characteristics; however, they are different.

| 4

> While bibliographic research fundamentally uses the contributions of several authors on a given subject, documentary research is based on materials that have not yet received analytical treatment, or that can be re-elaborated according to the research objectives. (PRODANOV; FREITAS, 2013, p. 55)

In this way, the documentary research was conducted by reading the documents of standards, guidelines, and manuals provided by IBICT, the body that maintains the BDTD. Regarding qualitative research, which "[...] considers a dynamic relationship between the real world and the subject, that is, an inseparable link between the objective world and the subject's subjectivity that cannot be translated into numbers. [...]" (PRODANOV; FREITAS, 2013, p. 70), allowed the analysis and interpretation of data in an elementary way to produce more in-depth information but without the need to collect statistical data.

While descriptive research, which according to Prodanov and Freitas (2013, p. 52), is used "[...] when the researcher only registers and describes the observed facts without interfering with them. It aims to describe the characteristics of a certain population or phenomenon or the establishment of relationships between variables. [...]" was based on the analysis of the retrieval systems of the bases of the repositories linked to the BDTD for the observation and description according to their functionalities and specifications.

The process of collecting and presenting research data took place in 3 stages. The first is the reading and interpretation of texts present in the literature that deals with the IRS as a concept, fundamentals, and structure, also about the IR and the information retrieval process in them; in this way, the proposal was fulfilled of the first specific objective of this research.

In the second stage, we sought to read the pages that address the concepts, purposes, and tools used in the BDTD project (IBICT page: http://wiki.ibict.br; BDTD portal: http://bdtd.ibict.br/, and the documents present in the database such as guidelines, manuals, and

metadata standards) and also the information made available in the systems of the participating repositories. With this, charts and figures are presented that summarize this information. In the third stage of this research, the 51 Higher Education Institutions (HEIs) repositories, a sample of this research, were observed, analyzing their interfaces and navigation and consultation methods for retrieving information in their bases.

The data analyzed in the research is presented in charts and figures, highlighting aspects to meet the objectives of this study. However, the information described is approached dialogically between the results retrieved according to the second and third stages of the research, not necessarily respecting the order in which they are inserted in this section.

3 INFORMATION RETRIEVAL

Information retrieval is related to the processes that involve access to information and its effectiveness in an appropriate way, depending on the informational demand of the user. This is done by retrieving documents with information relevant to the query and the information needs of the person seeking it. According to Cardoso (2004), it is an area that originates from computer science and is concerned with the storage and retrieval of documents.

Concerning information retrieval, Baeza-Yates, and Ribeiro Neto (2013) argue that this discipline focuses on the concern with the informational interests of users, seeking to promote easy access to information of interest. In this way, it is involved with the representation, storage, organization, and access to information. However, the actors describe that this concept is their initial objective, which has undergone a change, evolving beyond the indexing of texts and the search for relevant documents for users.

> Regarding research, the area can be studied from two distinct and complementary points of view: One centered on the computer and the other centered on the user. In the computer-centric view, [information retrieval] mainly consists of building efficient indexes, processing queries with high performance, and developing ranking algorithms to improve results. In the user-centered view, [information retrieval] mainly consists of studying user behavior, understanding their primary needs, and determining how this understanding affects the organization and operation of the retrieval system. (BAEZA-YATES; RIBEIRO NETO, 2013, p. 2)

In this process of access and use of information, both the system and the other tools that involve and make it possible are considered, acting on the development of mechanisms that work with the processing of data necessary for qualified results, adding value to the user's demand. As well as the user himself, seeking to know his behavior and his needs to better serve him and how, from this knowledge, he can be attributed to the system's operation.

When approaching the concept of information retrieval from the perspective of the views of different authors who deal with the subject, Oliván (2008 apud GODINHO, 2014) classifies them from two different perspectives, the first being composed of authors who conceptualize information retrieval related to the search process; and the second, being from authors who work with information retrieval regarding the search, as well as the representation and storage of information. However, he points out that between these two perspectives, there is a consensus in relating information retrieval to the concern of accessing information following users' interests.

The concern with the user's interests, regardless of the conceptual basis of ideas about information retrieval, is consistent with its reason for existing and is in constant development. The focus on the user contributes to the perception of the value of the service regarding the offer of information if considered in the process of construction and development of the service. "[...] the creation and delivery of value through products/services in an [Information Unit] will depend entirely on the user's perception. Thus, not all activities this organization performs will

| 5

have added value since this perception is centered on the service user." (INOMATA; PINTRO, 2017, p. 1009).

In the context of information retrieval in digital environments, this statement becomes valid since it will depend on a series of actions involving the system's functioning and the user's involvement with it when searching for an information service that satisfies his informational need. As the issue of adding value to information services is not a topic to be discussed in this work, it will not go into detail. Still, it is considered as a demand from the information system, which is a tool for accessing the information discussed here, is what was presented due to its connection with users and the purpose of having them as the focus of actions developed in the scenario of information retrieval.

Information retrieval involves processes that, in turn, include actions that consider the tools (for storage and retrieval) that innovate or improve it in the pursuit of qualifying the fulfillment of users' needs and how they interact and perceive the results. In the information retrieval process, according to Cardoso (2004), due to the large volume of information and the complexity of the types of data that are stored, there is a requirement that it be increasingly improved and, therefore, the information retrieval presents new challenges to be overcome, in addition to being characterized as an area with a high level of significance.

The need to seek to overcome new challenges gives the process different ways of working in its construct, including new means of retrieval and interaction with the systems and also improvements that contribute to facilitating access by the user, which directly affects the relevance of the results that provide a solution to the information needs.

3.1 Information Systems

Information has been treated as an indispensable resource in society's development, studied by different disciplines. And although recent in the field of science, it has come to be considered a tool directly linked to productivity (ARAÚJO, 2009).

Information acts in society through social actions, being present in all spaces, being taken as a condition in acts involving social development, serving as a determinant "[...] for the establishment of institutions, for the consolidation of research and for the sovereignty of nations, with direct repercussions on the way of life of individuals and their various practices." (SAMPAIO; LOUREIRO, 2019, p. 54).

As an example of this action of information in society and its spaces and spheres, it is possible to highlight the presence of information in the set of activities that involve applicability in other areas, understood as a primary input for scientific productivity processes (ARAÚJO, 2009).

"Considering the information problem as defined, that is, the *informational explosion*, information retrieval became a successful solution found by [information science] and in the process of development until today [...]" (SARACEVIC, 1996, p. 44, emphasis added). For information retrieval, it is necessary to use systems that facilitate access to the information after the information is indexed in a base in which the system performs the search based on operations that involve the terms and methods used by the users, from the use of search engines that are available in the interfaces of information systems.

For Baeza-Yates and Ribeiro Neto (2013), IRS can be classified into three categories of information retrieval models: Based on texts, based on links, and those based on multimedia objects (Figure 1).

Figure 1. Taxonomy of IR Models.
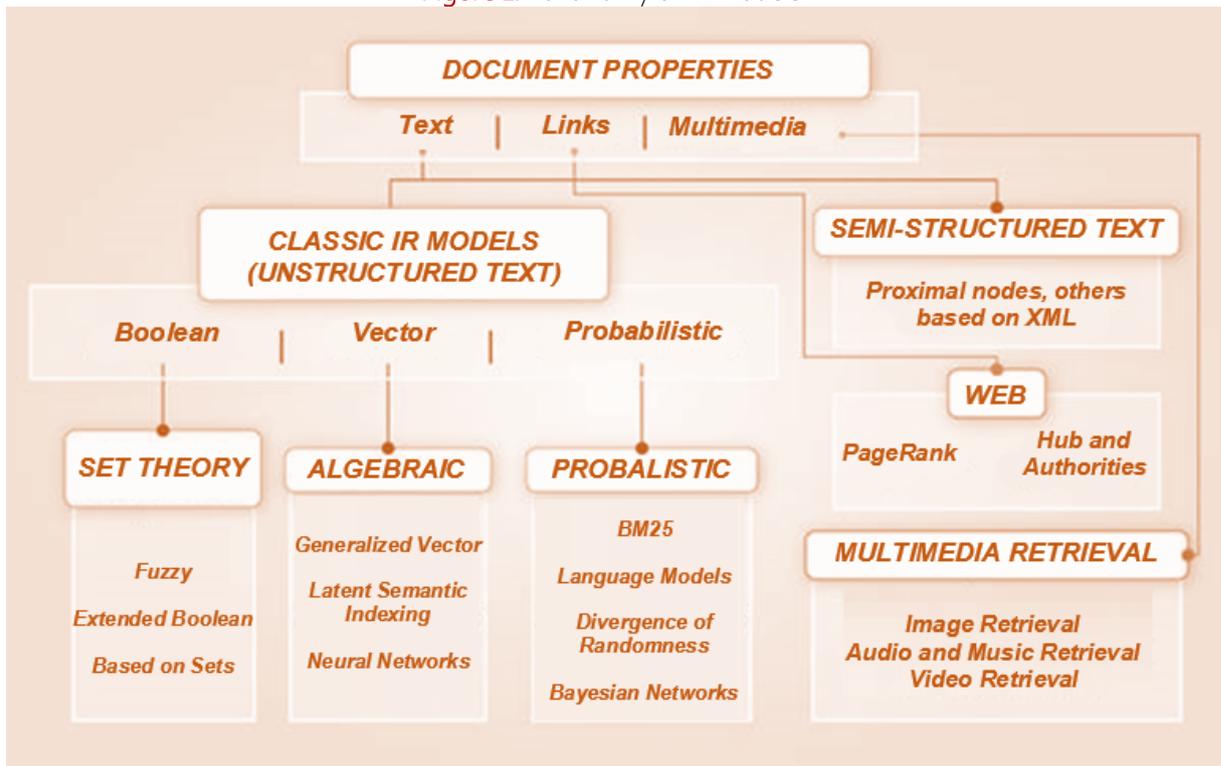
Source: Adapted from Baeza-Yates and Ribeiro Neto (2013, p. 24) by Monteiro et al. (2017, p. 164).

The first constitutes the basis of the IRS construct, and the others present specificities related to different types of search, documents available on the Web, and multimedia files. Documents indexed in a database have different typologies and, as such, require particular ways of searching, as shown in Figure 1. Souza (2006) understands IRS as systems that:

> [...] organize and facilitate access to information items, performing the activities of representation of information contained in documents [...]; Physical and/or logical storage and management of these documents and their pictures; Retrieval of the represented information and of the stored records to satisfy the information needs of the users [...]

Still, according to Souza (2006), this need is characterized as a timeless phenomenon, which is the basis for the concepts and definitions established by theorists highlighted by the author, such as Araújo (1995), Lancaster and Warner (1993); Lancaster (1968, 1979); Baeza-Yates and Ribeiro Neto (2013), among others, who approach the theme from the different methodologies and technologies used to meet the information needs of users.

For Araújo (2012), the functions configuring the documentary communication process must be established from the relationship between an IRS and the user. Therefore, the system's primary function that aims to information retrieval is to have information from the indexed documents, synthetically describing them with a clear and objective representation. Considering its various processes, the author highlights selection, acquisition, indexing, search, and retrieval as part of its actions.

It is noted that the IRS involves processes, and such processes have demands linked to information and its indexing (updating and availability); users (skills in interacting with the system and their informational needs); and a set of communication and information technology tools – ICTs, which enable better interaction between the user and the system, to retrieve information that corresponds to the informational demands of the user.

Based on these considerations about the IRS, the following subsection deals with the processes that involve the information retrieval in institutional repositories, highlighted in this research as a means by which the user, mainly as a social agent, can find solutions to their

| 7

questions, and as a place that gives the value based on visibility to the scientific productions of postgraduate program researchers, precisely theses and dissertations.

3.2 Open Science Movement

The Open Science Movement refers to a set of practices and principles that seek broad access to scientific knowledge in an accessible, shared, and reusable way. One of the main objectives refers to making research and scientific data accessible to all, seeking to promote transparency and collaboration in the scientific community. This movement encompasses a series of practices aimed at democratizing scientific knowledge (UNESCO, 2023).

Its efforts focus on increasing transparency, collaboration, and public participation in scientific research. By opening data, research results, methods, and publications, we seek to promote the dissemination of studies free of cost, allowing other researchers to verify and build on existing developments.

According to Abagli (2015), there is a movement with a global reach that reveals the inadequacy of the predominant methods of scientific production and communication. These methods are affected by artificial mechanisms that create obstacles of different natures, mainly in the legal and economic spheres, limiting their free circulation, collaboration, progress, and dissemination. Surprisingly, despite few technical barriers to the instantaneous circulation of information, this movement acquired a significant international dimension.

In this scenario, scientific communication, amid the use of digital tools, has undergone a paradigm shift, with open access being considered as the new model of communicating science, which was initiated from the action of events in favor of the Open Access Movement as the periodicals crisis that occurred in the 90s due to the high prices of scientific periodicals, subscribed by libraries to have access to information, and the Santa Fe Convention (1999), which culminated in the Open Archives movement (OIA) (WEITZEL, 2005).

There are also several statements, of which we can mention the three main ones highlighted by (ANDRADE; MURIEL-TORRADO, 2017): Budapest Open Access Initiative (OAI, 2002; Bethesda Declaration on Open Access Publication (2003); and the Berlin Declaration on Open Access to Knowledge in the Sciences and Humanities, both concerned with dissemination and implementation actions in favor of open access to information.

At a national level, one of the first discussions regarding open access to knowledge took place based on ideas on the subject in universities, followed by the incentive by the IBICT for the "implementation of tools" based on the Open Access Movement, being the cause of the emergence of several studies, and their discussion in events, in addition to the implementation of several tools (SANTOS JUNIO, 2010). Among the first dialogues with the movement, IBICT was one of the forerunners, "[...] which, through pioneering initiatives, has enabled [Brazil] to rise to the level as one of the nations that most [support] the Open Access Movement, through the development and implementation of tools that provide such end [...] (SANTOS JUNIOR, 2010, p. 32).

According to Weitzel (2019, p. 108-109),

> among the actions implemented, the translation and training of the leading digital platforms for publishing journals and conference proceedings stand out, as well as the implementation of repositories – respectively, the Open Journal System (translated to Portuguese as an *Sistema Eletrônico de Editoração de Revistas – SEER* with more than 1,500 journal titles); Open Conference System (translated into Portuguese as the *Sistema Online de Acompanhamento de Conferências – SOAC)*; and [DSpace] – software adopted for implementing repositories.

Implementing repositories such as the Open Journal System (SEER) and the Open Conference System (SOAC), together with DSpace software, enables better organization, preservation, and sharing of educational materials, enriching the research environment, and

fostering scientific progress collaboratively. These actions reflect an advanced vision aligned with the current demands of the academic community.

A national initiative that meets the greenway strategies refers to the development of the BDTD Portal, which collects the descriptions of theses and dissertations from the repositories of HEIs and research centers, making them available in a single portal with national visibility, which favors access and use of studies produced by researchers from these institutions (IBICT, [2022]), also developed by IBICT, together with the collaboration of the National Council for Scientific and Technological Development (CNPq) and other teaching and research institutions, which are mentioned in section 4.1, along with further descriptions about the BDTD, its origin, objectives, characteristics, and functionalities.

3.3 Information retrieval in Institutional Repositories

The promotion of the Open Access Movement is characterized by two strategic ways, the golden way, referring to the publication of articles in open access journals, and the green way, which deals with the self-archiving of scientific productions in open access. These two main channels contribute to its wide dissemination and the strengthening of its objective, democratizing access to information. The green way (Green Open Access), the focus strategy of this research, characterizes that the repositories that contribute to open access must be institutional.

> An open access institutional repository constitutes a scientific information service - in a digital and interoperable environment - dedicated to managing an institution's scientific and/or academic production (universities or research institutes). It contemplates the gathering, storage, organization, preservation, **retrieval**, and, above all, the wide dissemination of the scientific information produced in the institution. (LEITE, *et al.*, 2012, p. 7, emphasis added)

Information retrieval generally involves processes that demand a series of tools that collaborate so that its primary objective, meeting the informational needs of users, is achieved. For information to be disseminated and consequently used, it is essential to communicate it. Scientific communication uses means to reach its users broadly. The use of digital media has been helpful in this action. As such, it has been bringing about changes in publishing and disseminating science. Digital repositories emerged in this scenario of paradigm shifts in scientific communication, serving as a channel for scientific dissemination (CAFÉ; MUÑOS; LEITE, 2015).

It is worth mentioning that IR differs from digital ones. Thus, all IR is digital. However, not all digital repositories are institutional. In a broader context, "Digital repositories are created to facilitate access to scientific production. These are databases developed to gather, organize, and make the scientific production of researchers more accessible." (LEITE, *et al.*, 2012, p 7). However, they are not necessarily managed by institutional bodies. While the IR has the characteristic of being institutional for this reason, dealing with the storage, preservation, dissemination, access, and use of scientific productions produced in the academic field.

According to Café, Muños, and Leite (2015, p. 33),

> institutional repositories are critical scientific information systems, as they facilitate communication between researchers, guarantee free access to scientific knowledge, preserve an institution's scientific production in an organized manner and protect researchers from plagiarism. [...]

Since IR are information systems, they have the demand to better meet users' needs. However, it must be inferred that the ways of using the system predispose to competencies in the act of consultation carried out by the user to obtain information relevant to their informational questions. One of the difficulties concerning relevance or not, raised by Souza

| 9

(2006, p. 164-165), which he considers a central problem in information retrieval, deals with the

> [...] prediction of which documents are relevant and which must be discarded, and this 'choice' task, in automated systems, is performed by some kind of algorithm that, based on previously defined heuristics, decides which are the relevant documents to be retrieved and orders them based on established criteria [...]

These functions and usability will depend on the IRS model, knowing there will be consultation processes (user action) and data processing (by the system). In this way, this work is based on the idea that, in addition to defending the open access movement, understanding the processes that involve the information retrieval in open access IR systems is configured as a demand for achieving visibility and the effective use by users who seek solutions that meet their informational questions.

The information retrieval in IR needs to understand the digital resources provided by the systems, such as their search engines, which are directed to provide the best visualization and access to information. Considering that the IRS have interfaces that allow users to translate their information needs through questions or keywords and examine the documents contained in the repository to find information relevant to their search (SOUZA, 2006).

Based on the discussions presented above, the section that follows, as well as its assigned subsections, deals with the analysis of the IRS from the visualization of the search tools of the Brazilian institutional repositories of theses and dissertations available in the BDTD and its databases. Because of this, we sought to inquire about its characteristics related to the retrieval of theses and dissertations that make up and structure its databases.

4 RESULTS AND DISCUSSIONS

This section presents the results and discussions regarding the analyzes carried out according to the methodology applied in this research.

| **10**

4.1 Brazilian Theses and Dissertations Repositories: Analysis of Search Engines

BDTD, officially launched in 2002, was created and is maintained by the Brazilian Institute of Information in Science and Technology (IBICT). Its creation took place from the Brazilian Digital Library Program (BDB), financed through the support of the Financier of Studies and Research (FINEP). The construction project was developed based on the creation of a technical advisory committee (CTC) (BDTD, [2022]).

The committee was formed by representatives of IBICT, CNPq, Ministry of Education (MEC) represented by the Coordination for the Improvement of Higher Education Personnel (CAPES), Higher Education Secretariat (SESu), FINEP, University of São Paulo (USP), Pontifical Catholic University of Rio de Janeiro (PUC-RIO) and the Federal University of Santa Catarina (UFSC) (BDTD, [2022]). Thus, the BDTD now has 7 main lines of action that would culminate in its implementation (Chart 1).

Chart 1. Lines of Action of the BDTD.

Lines of action of the BDTD
Study existing experiences in Brazil and abroad in the development of digital libraries of theses and dissertations;
Develop, in cooperation with community members, a model for the system;
Define metadata standards and technologies to be used by the system;
Absorb and adapt the technologies to be used in the implementation of the model;

Develop an electronic publication system for theses and dissertations to serve those teaching and research institutions that did not have automated systems to implement their digital libraries;

Disseminate the adopted standards and technologies and provide technical assistance to potential partners in their implementation;

Study existing experiences in Brazil and abroad in the development of digital libraries of theses and dissertations.

Source: Adapted from bdtd.IBICT.br ([2023]).

Since 2003, BDTD has significantly changed its system structure and metadata standards. Currently (2022), the Library has technological tools, metadata standards, and management systems, which are made available to institutions wishing to develop their repositories and be part of the BDTD (Chart 2).

Chart 2. Tools for structuring and developing institutional repositories.

Item/Tool	Functionality
Electronic Publishing System of Theses and Dissertations (TEDE2) (in its second version/update)	"The TEDE System allows the automation of electronic publication procedures for theses and dissertations (TDEs), cooperatively involving the main actors in the publication process of a thesis or dissertation, such as authors, postgraduate programs, and libraries." (TEDE UFC, [2022]). It can be used for the creation, management, and effectiveness of digital libraries of each institution and electronic publication of their theses and dissertations.
Brazilian Standard Metadata for Theses and Dissertations (MTD3-BR) v. 3 (in its third version/update)	Set of metadata used for describing theses and dissertations, aiming to ensure interoperability between information systems (BDTD and systems of participating institutions), facilitating data collection and control. This is based on Dublin Core standards, adapted for Brazilian metadata.
OAI-PMH Protocol Layer	The standard accompanying the TEDE system aims to guarantee the interoperability of the individual systems of the participating libraries.
Dublin Core Standard	The standard system for describing digital objects (theses and dissertations)
Manuals	Procedures for installing, connecting, and running the system.

Source: Research data.

According to the consultation carried out in the National Register of Courses and Higher Education Institute (e-MEC) on August 25, 2022, based on filters in the search engine of the portal system (search for Higher Education Institute; Administrative Category: Federal Public and University; Status: Active), 68 academic organizations (universities) are of federal public admiration status.

To situate the research investigation sample, the IR of the federal HEIs, a query was carried out on the BDTD portal, listing the institutions listed by e-MEC. It was found that of the 68 HEIs consulted in the National Register, 51 are BDTD participants, which have IR developed according to the metadata standards established by the BDTD guidelines, being possible to collect and make their data available in the Digital Library (APPENDIX A).

Given the sample, the information related to the observed and analyzed information of the IRS of the repositories is presented below. An attempt was made to describe the metadata standards provided by the IBICT and those contained in the individual libraries of each institution. And also about structuring the interface and its search engines, highlighting the ways to search, search resources, how the results are presented, visual elements structured in the

TEDE2 interface, and issues related to interface customization. And also how to carry out the process of collecting data from the IR for the national portal, the BDTD.

4.2 Standards and Guidelines Established by IBICT

For its functionality and interoperability, IBICT and the technical committee approved a set of metadata to perform the description of theses and dissertations, the Brazilian Metadata Standard for Theses and Dissertations (MTD-BR). The Standard is specific to the description of theses and dissertations. According to the BDTD Portal ([2022]),

> the use of the standards and guidelines of the Brazilian Digital Library of Theses and Dissertations (BDTD) is essential to ensure interoperability between information systems. Since its inception, the BDTD has used the Brazilian Metadata Standard for Theses and Dissertations (MTD-BR). The Brazilian Standard was also developed by analyzing other international standards for describing this document typology, such as the Electronic Thesis and Dissertations Metadata Standard (ETD-MS). The Brazilian Standard has also been updated over the years following changes in international standards, new systems, and formats. Thus, today, the BDTD uses the third version of the Standard, the MTD3-BR.

The updates aimed at improving the control and quality of the description of the data collected and handled electronically; the first version, MTD-BR, is made up of 71 metadata, while the second version, MTD-BR2, is made up of 80 metadata (BAPTISTA; FERNEDA, 2016). And the current version, MTD-BR3, was reduced, consisting of 38 metadata, and developed based on the Dublin Core metadata standard, adapted for the Brazilian reality (TEDE PUCRS, [2022]).

The standards are further divided into mandatory and optional metadata; the mandatory aim is to maintain interoperability between the systems so that the BDTD can collect theses and dissertations of the IR and make them available in the National Digital Library (Table 1).

| 12

Table 1. Versions of the MTR-BR Metadata Standard.

Versions of the (MTR-BR Metadata Standard)	Year of Creation	Required Metadata	Optional Metadata
Brazilian Metadata Standard for Theses and Dissertations Description (MTD-BR) v. 1	2002	38	33
Brazilian Metadata Standard for Theses and Dissertations Description (MTD-BR2) v. 2	2005	31	49
Brazilian Metadata Standard for Theses and Dissertations Description (MTD-BR3) v. 3	2012	18	20

Source: Research data.

The development of the MTD-BR Standard was an integral part of the BDTD implementation project, intending to generate information products and services whose function refers to identifying and locating electronic theses and dissertations. Furthermore, it allows the collecting of information that generates indicators and integration with other national repositories of teaching and research information in Brazil (LOURENÇO, 2005). The current Standard, MTD-BR3, comprises 28 metadata, 18 mandatories intending to ensure operability between systems, and 20 optional metadata.

4.3 Interfaces and Search Engines of Institutional Repositories Retrieval Systems

As previously mentioned, BDTD is established through technological tools, such as software, protocols, and data description standards. Its environment is cooperative, consisting of three components, the local BDTD (from which theses and dissertations are collected, many of which have the TEDE System), the collector software, and the national BDTD that brings together the productions arranged in the local BDTD. Data are collected via the OAI-PMH protocol, enabling the systems' interoperability based on the mandatory metadata set and using XML technology (IBICT, 2014).

For the development of IR, IBICT developed the TEDE system,

> This system is an application developed in a database environment, implemented in HTML, and geared towards the WEB. It features a user-friendly interface with a menu-driven navigation system. It has a toolbar for navigating the system's operation modules; it has automatic help in the existing Help option in the menus of the screens for running the operation options of each module; it has operating commands for implementing the actions required throughout the electronic publication process; it has Choice Boxes for selecting data contained in tables provided by the system and command buttons for confirming the actions performed. (TEDE UFC, [2022]).

The second version of TEDE is based on the DSpace[1] software, updated to follow the evolution of the system's development since the first version is considered obsolete, equipped with customizations that adapt to the needs that deal with these and dissertations, having configuration specifications, layout modifications, and development items. With DSpace, it is possible to manage and preserve digital objects and facilitate the process of their retrieval. DSpace has several purposes, while the software is used to implement repositories following free access, thus agreeing with open access (IBICT, 2014).

Still, according to IBICT (2014), DSpace for this purpose is highly configurable and has evolved in improvements and ease of use from innovations that arise according to the needs of its user community, such as full-text indexing, the use of controlled vocabularies and the XMLUI interface. And yet, considered one of the most significant changes made to DSpace, the possibility of using two Web interfaces (JSPUI - Java Server Pages User Interface; and XMLUI - eXtented Mark Language User Interface), which generates the opportunity to choose between two interfaces that have different technologies (IBICT, 2014) (Figure 2).

| 13

Figure 2. DSpace Infrastructure.
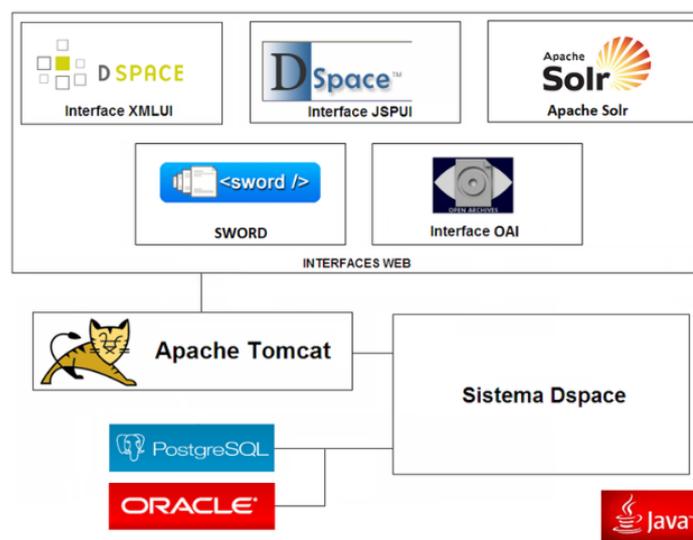
Source: IBICT (2014).

---

1 Available at: https://dspace.lyrasis.org/. Accessed on: Feb 4, 2023.

In this way, the decision of which interface to use rests with the administrator or the repository management team. Regardless of the choice, there will be no system operability problems, as there is only one database and one set of programs. TEDE2 is customized in the JSPUI interface of DSpace and sought the appropriate way to store information from theses and dissertations and guarantee the availability of descriptions to be collected by the national BDTD (IBICT, 2014).

4.3.1 Search and Navigation Engines

In general, the repositories follow the DSpace standard; however, in the case of TEDE2, it is possible to observe implementations regarding access to documents based on the interface, as specified above. The institutions participating in the BDTD have a specific repository for storing the thesis and dissertation document typology (UFMA, UFAM, UFRPE, and UFRRJ), use a community composed of collections of theses and dissertations, or in general, configure them in the same environment, where you can specify the type of documents by theses or by dissertations in the faceted search.

For analysis and description of the data referring to the forms of navigation and retrieval of information in the SRI of the repositories, the reading and interpretation of the information available in the tab "Ajuda/Help/?" standard for repository systems was conducted. Using as a basis the UFMA Digital Library of Theses and Dissertation (referring to the navigation modes in TEDE2) and the UFPB IR (which uses DSpace in its XMLUI interface).

According to the definitions analyzed in the help tab accessed through the UFPB Repository, which uses the DSpace XMLUI interface. DSpace has an interface for viewing and browsing data from digital documents, including theses and dissertations, the type of material TEDE2 fits. Articles, working documents, such as TCC, theses and dissertations, preprints, technical reports, conference documents, and data sets in various digital formats can be managed.

In this way, the user navigates through a specific list of items that fit to browse by Community or Collection, title, author, subject, and date. The system also contains a navigation bar, allowing simple or advanced queries using the Jakarta Lucene search engine. Researched terms will be searched in the fields title, author, summary of the subject, series, sponsor, and identifier of the record of each item (REPOSITORIO UFBP, [2022]).

The user can use research methods such as truncation, stemming, research by phrase using an asterisk (*) after the root of a word, the plus sign (+) to research for exact words, the use of double quotes ("") to research for the complete set of words or phrases and Boolean features (AND, OR and/or NOT). As well as researching by subject category using controlled vocabulary, which deals with a set of terms for describing subjects or themes related to a specific area (REPOSITORIO UFBP, [2022]).

For the conceptualization and description of the forms of navigation and research mentioned above, the navigation and consultation methods present in TEDE2 are described below, based on the DSpace software, however, adapted for the type of documents dealing with electronic thesis and dissertations. TEDE2 is customized based on DSpace's JSPUI interface and has specific fields for the type of document for which it is intended.

The system's content using DSpace is structured hierarchically. It may contain Communities organized by administrative entities depending on the institution (departments, thematic areas, type of document, etc.) (Chart 3). An unlimited number of subcommunities can be added to Communities, which represent the highest level in the hierarchy. And both are formed digital documents (called items), which are grouped in Collections and can also contain an unlimited number of collections (TEDEBC UFMA, [2022]).

| 14

Chart 3. Navigation and query in DSpace JSPUI interface.

Browse (see items in a specific order)	Simple Research	Advanced Research	Boolean Operators	Search
By Community and Collection allows you list Communities in alphabetical order. A slightly indented list of its subcommunities and Collections will appear for each Community. By Date allows you to view a list of items in chronological order. By Author allows you to see a list of items sorted by author's name. By Title allows you to list, in alphabetical order, the titles of items from all Collections in the system. An auxiliary search tool will appear at the top to help locate the desired item. By Subject allows you to view items by an alphabetical list of subjects.	To research TEDE, you can use the simple research tool at the top of the navigation menu or the research in the middle of the main page. The research result will offer a new form where you can narrow your research to a specific Collection or Community. The Jakarta Lucene system (High-Performance research Engine) is used. The research is done on the item's metadata (on full text depending on system configuration), such as author, title, etc. The word(s) used for the research are then compared with those existing in the fields of each item. The Stop Word is used (words such as articles, prepositions, and conjunctions are not used in the research). The user can resort to truncation by using an asterisk (*) after a word fragment, indicating that he wants to research for items with words beginning with a word fragment. Morphological variations can be retrieved by	Advanced research allows you to specify fields and make combinations between fields with the logical operators "AND (AND)," "OR (OR)," or "NOT." The research can be restricted to a community or the entire system. To restrict the research to a community, select it from the list at the top of the form. The desired field can be selected from the list on the left and enter the term on the right. If you want to use multiple fields, choose the logical operator between fields and the field, and enter the research term. You can restrict the research through a faceted search by choosing one of the specified items (author, supervisor, committee member, all contributors, CNPq area, type of document, and year of defense).	AND (to group terms and return documents that contain both terms grouped together). OR (to return documents with the queried terms, choose one or both present in the same document). NOT (to exclude a particular term, not returning documents that have it). You can make use of Boolean operators together. For this, parentheses can be used to group search terms, valid in more complex researches that will require using more than one of the operators.	By Subject Category, you can filter the list of categories, removing all terms that do not match the filter. The remaining terms are any category or subcategory that contains the filter term anywhere in the hierarchy. To research the archived items by subject category, check the boxes next to the categories as needed. Clicking the "+" sign next to the category will expand the tree to show the available refinements for the selected category.

| 15

(Stemming): The search engine automatically expands the word and can retrieve plural or inflected words.
You can research by phrase: To research for compound terms or phrases, put them between double quotes ("). The system will search for the expression or phrase precisely as written.
You can research precisely by putting a plus sign (+) before the desired word in the result. Or, eliminate items with unwanted words by adding a minus sign (-) before the word that should not appear in the research result.

Source: Data taken and adapted from tedebc.ufma.br ([2022]).

Analyzing the systems that use TEDE2, it was found that the interfaces present the navigation and query specifications as detailed in Chart 3. In this way, the search type was observed, the possibilities of using search operators, how the results are presented, and visual resources that can be accessed and/or customizations present (Chart 4).

Chart 4. Navigation in IR implemented with TEDE2.

Search types	Search operators	Results presentation	Visibility/ Customization
Browse communities and collections (the collection of theses and dissertations specific to each graduate program). Browse by: Defense date, author, supervisor, title, subject, CNPq areas, department, Graduate Program, type of document, and type of access. Search in: The entire repository or the institution's university fields.	Boolean operators (AND, OR, and NOT). Proximity operators manually (quotes and parentheses)	Results and pages (5-100) Sort by: Relevance, defense date, title, author, supervisor, program, and type of document. Sort (ascending and descending). Record: All or from 1 to 50. Displayed list of documents, according to	Language: English, Spanish, and Portuguese. Word cloud with subjects with the highest number of records (by clicking on m of the terms, the system directs to a list of documents related to the specified subject).

Simple Search (direct text in the research bar); and Advanced (application of filters and operators and search saving current terms, specifying by type of document, title, author, supervisor, all contributors, contributor identifier, summary or abstract, subject, CNPq area, date of defense, date of release for access, type of access, institution, initials of the institution, program, department, development agency, file format and all fields), and determining how the search terms will be researched: Equal, contains, identified, different, does not contain and unidentified).

the ordinance established by the user, with the items: Date of defense, title, author, program, and type of document.

Faceted by: Author, supervisor, committee member, all contributors, CNPq areas, type of document, and year of defense.

Source: Research data.

The interface in the TEDE2 System can be presented by the IRS of the HEI repositories according to the system specification. Still, they can also be customized according to the institution's criteria. In the analysis of the repositories, only a few changes were found. Some add all available items, while others use a portion of them. For example, when using TEDE2, 17 filters are available containing items to be specified in the search. However, the libraries that opted for TEDE2 use only 7 of the faceted search items. On the other hand, they all show the visualization using a word cloud highlighting the main subjects, according to the TEDE2 interface (Figure 3).

| 17

Figure 3. TEDE2 System Interface.

Source: IBICT (2014).

Just below the word cloud, filters called Faceted Search are presented. At the top, users can refer to other information related to navigation or related to the system. In TEDE2, the

fields are presented: *página inicial* [homepage], *Navegar por* [Browse by], *Ajuda* [Help], *Sobre o TEDE* [About TEDE], and *Entrar* [Sign in], as shown in Figure 4.

Figure 4. Top tab of the TEDE2 interface.

On the '*página inicial* [homepage],' the user returns to TEDE's main page, which can be executed by clicking on the institution's logo. As for '*Navegar por* [Browse by],' you can use the system's navigation tools and select to access by Date of defense, Author, Advisor, Title, Subject, CNPq Areas, Department, Graduate Program, Type of document, and Access type.

In the '*Ajuda* [Help]' tab, you can check information regarding the process of collecting information from the system (about Harvesting), the most frequent questions from users and their solutions (Frequently asked questions), and concerning the repository, the forms of navigation, its structure, and organization, as well as instructions for using search engines and specifications for researching and navigating the system (Standard Help).

In '*Sobre o TEDE* [About TEDE],' the user will have information that addresses the processes of creation, management, and use of the TEDE system. While in '*Sobre a BDTD* [About BDTD],' information consistent with the processes related to BDTD is accessed. By clicking on '*Entrar* [Sign in],' the user can register or log in with the platform in '*Meu espaço* [My space],' activate resources for receiving updates in '*Receber atualizações por e-mail* [Receive updates by email],' and make changes to his profile in '*Editar perfil* [Edit profile].' At the bottom, information about the institution, such as its logos and other relevant information, can be added as a footer.

The definitions specified in TEDE2 seek to facilitate access to a specific type of document, theses, and dissertations; thus, the observed resources provide easy access to the content available in the repository, contributing to the research needs of users who are part of its community.

Institutions that do not use TEDE2 in interface systematization and navigation methods are also based on DSpace standards. The differences are in the definitions specified from the choice of one of the interfaces provided by DSpace, or the attributes used from them, which can also be customized, as in the case of TEDE2, which was developed based on DSpace JSPUI.

4.4 IR System Customization

Institutions participating in the BDTD network have access to a data package containing the TEDE system, the OAI-PMH protocol, and manuals with installation and use instructions. The TEDE system was developed so that libraries can automate their IR, however,

> The BDTD can be integrated without Brazilian teaching and research institutions using a single software to manage theses and dissertations. However, the system must export data from theses and dissertations in XML format so that, through a converter, IBICT can automatically collect (harvesting) the metadata. Also, the system can directly export documents in the metadata standard defined by the BDTD, subject to filling in at least the mandatory fields. (TEDE PUCRS, [2022])

In this way, libraries can either use other software or TEDE2 as long as it uses the 18 mandatory metadata of the MTR-BR3. And so contribute to collecting the data made available

| 18

in their systems. Institutions can also, using TEDE2, customize the software interface and define the workflow in the system (BDTD, [2022]).

The customizations found in libraries are attributes that give greater visibility to the data contained in the IR. They are tools linked to the interface, adding value to the documents, such as word clouds with the main subjects addressed in theses and dissertations produced in the institution, as in the case of the Repositories that use TEDE2, miniatures of documents in the results list, table with the last documents deposited in the base, among others.

It is verified that that the institutions participating in the BDTD do not necessarily need to stick to the standards in a partial way established by IBICT, being necessary only that they assign to their repositories the mandatory metadata standards, along with the use of the OAI-PMH protocol and the use of XML technology so that the collection of data occurs for availability in the national BDTD portal.

4.5 Process for Collection and Availability of Theses and Dissertations on the BDTD Portal

In the use of its functions and operability, BDTD uses free solutions. The same occurs in collecting theses and dissertations in the IR of the participating institutions and making them available on the BDTD portal. BDTD "[...] uses the free software VuFind to aggregate and make available theses and dissertations collected from the information systems managed by the institutions. Recently, BDTD started to adopt the collector developed and used by **Red de Repositorios de Acceso Abierto a la Ciencia (LA Referencia)**" (BDTD, [2022], emphasis added).

Through a technical team, the institution La Referencia developed the Free Software Platform (GPL3), which collects informational data and is used by member countries, including Brazil, as an aggregator service and national portal. The platform was created based on the free software VunFind, consisting of three components that work together (LRHarvester, LRProvider, and Search Engine/Service Portal). Thus, each component is responsible for carrying out data collection in a coordinated manner (Chart 5).

| 19

Chart 5. GLP3 Components.

Components	Responsibilities
LRHarvester	"[...] OAI-PMH metadata collection, transformation, and validation component. Responsible for collecting OAI sources, validating metadata records, and transforming them to meet network quality guidelines. The collection component was initially based on OCLCHarvester2. The platform is developed in Java 1.7 language, using Spring framework and PostgreSQL 9.X database engine. It also works over MySQL." (LA REFERENCIA, [2022]).
LRProvider	"[...] component responsible for publishing metadata using the OAI-PMH protocol, currently derived from the Dspace-XOAI software, which offers high flexibility and scalability for interoperability with other international metadata aggregators." (LA REFERENCIA, [2022]).
Search Engine/Service Portal	"[...] aggregated metadata publishing portal that provides a user-friendly and mobile-friendly research interface. It also provides statistics on metadata quality and crop evolution. This component is based on the free software Vufind (http://vufind.org) on which specific modules have been developed." (LA REFERENCIA, [2022]).

Source: Adapted from La Referencia ([2022]).

The process starts with collecting information in the IR systems, followed by transforming and validating the OAI-PMH protocol metadata used on the BDTD platforms (*LRHarveste*). With the recovered data, the metadata is published (*LRProvider*) from the metadata portal with a friendly interface (Search Engine/Service Portal), being made available

on the BDTD, which gives visibility internationally to the theses and dissertations produced at the national level in institutions participating in the BDTD.

Collection platform design principles were defined in 2014 and 2015 and still prevail. Altogether, ten characteristics define and establish the GPL3, making it usable in collecting information: Transferable, responsive, multilingual, business oriented, support and flexibility, scalable, manageable, and friendly, possibility of manipulation, Open Source, statistics, and economic (Chart 6).

Chart 6. Open source platform design principles.

Design Principles (2014-2015)
1 Transferable: The version of the central node must be the same as that of the national nodes.
2 Responsive: Access via mobile devices (bootstrap).
3 Multilingual: English and Portuguese support.
4 Business-oriented: Focus on OAI metadata validations and transformations.
5 Support for OpenAIRE guidelines and flexibility to update new guidelines.
6 Scalable: It should collect hundreds of deposits and over a million records in less than 24 hours.
7 Manageable and user-friendly: Diagnostic, validation, and reporting tools for administrators. Possibility of handling by information specialists and not just by IT technicians
8 Open Source: Development and use of components and solutions (for example: VuFind for the front-end or use popular databases like MySQL, PostGres, and Java).
9 Statistics: Development of new modules
10 Economical: Based on free Linux distributions such as Ubuntu and a 'medium' national node, a dual-core machine with 8GB RAM and 500GB space.

Source: Adapted from La Referencia([2022]).

According to information from BDTD ([2022]),

> BDTD uses the technologies of the Open Archives Initiative (OAI), adopting a model based on interoperability standards consolidated in a distributed network of several information systems that store theses and dissertations in their databases. This model has the fundamental presence of two main actors: The data provider and the service provider.

| 20

Thus, two main agents act in the implementation of the model, the data provider, which are the participating institutions, and the service provider, the role played by IBICT. Thus, IBICT "[...] manages the deposit by exposing the metadata and the full text for automatic collection (harvesting)" (BDTD, [2022]). While institutions "[...] provide information services based on metadata collected from data providers." (BDTD, [2022]). These actions focus on achieving the main objective of the BDTD, which is to gather and make available in a single portal the theses and dissertations produced and defended by Brazilian researchers working in Brazilian teaching and research institutions and outside the country.

5 CONCLUSION

In the context of the discussions highlighted in this work, it is verified that information retrieval involves processes attributed to access to knowledge produced in teaching and research centers and that the technological resources that can be used to facilitate access to information relevant to an informational demand, are aligned with the main objectives of the movement in favor of open access, as, from the use of repositories for the search and access to information, they confer both concerns regarding computational tools and the demands of users.

Observing the interiors of the repositories, such as structure, characteristics, and available tools, it was identified that the information retrieval systems of the repositories have technological tools and specifications that confer navigation in the systems by the users,

contributing to the best search experience and access to information relevant to their informational issues.

The forms of navigation and access seek to serve the user, using tools that allow the information retrieval consistent with the needs of its users; in addition, it contributes to open access by making information related to documents available in a single portal, which points to the access of the document in the IR database, making it possible to read it on the page or download the file.

When comparing the search tools of the systems with the parameters established by the IBICT, it can be observed that the information systems of the repositories meet the requirements concerning the search forms identified by the parameters of the BDTD; the main differences are in the customizations verified in each repository, (as the attribution of visual resources: Word clouds, book cover icons in the presentation of search results, among others).

IBICT, when developing a tool and worrying about updating it according to the needs of its community, through the use of digital tools that provide easy access to relevant information. In addition, dealing with a documentary typology, electronic theses, and dissertations, the TEDE2 system has resources that favor their retrieval, shaping the available tools according to the dispositions of theses and dissertations, in addition to adapting high-standard international technologies to reality Brazilian, without changing the standards that make it possible to make national scientific productions available in international databases.

Based on the current discussions, it is understood that future research can work on the main reasons for the absence of teaching and research institutions that are not participants in the BDTD, based on an analysis of the main advantages of their participation, through the development and management of institutional repositories and consequently the availability of the scientific productions of its agents on the BDTD portal.

REFERENCES

ALBAGLI, Sarita. Ciência aberta em questão. *In*: Albagli, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Cap. 1, p. 9-25. Available at: https://9x0o.short.gy/questõesabertas. Access: 10 jun. 2023.

ANDRADE, Rebeca de Moura; MURIEL-TORRADO, Enrique. Declarações de acesso aberto e a lei de direitos autorais brasileira. **Rev. Eletron. Comun. Inf. Inov. Saúde**, Rio de Janeiro, v. 11, nov. 2017. Available at: https://www.arca.fiocruz.br/bitstream/icict/23722/2/30.pdf. Access: 22 jul. 2022.

ARAÚJO, Carlos Alberto Ávila. Correntes teóricas da ciência da informação. **Ci. Inf.**, Brasília, DF, v. 38, n. 3, p.192-204, sep./dec., 2009. Available at: http://bit.ly/3HyUwqP. Access: 19 jun. 2022.

ARAÚJO, Vera Maria Araújo Pigozzi de. Sistemas de recuperação da informação: uma discussão a partir de parâmetros enunciativos. **Transinformação**, Campinas, v. 24, n. 2, p. 137-143, 2012. Available at: https://brapci.inf.br/index.php/res/v/116379. Access: 26 dec. 2022.

MÁRDERO ARELLANO, Miguel Ánge; LEITE, Fernando César Lima. Acesso aberto à informação científica e o problema da preservação digital. **Biblios**, Brasília, n. 35, mar./jun., 2009. Available at: https://repositorio.unb.br/handle/10482/4937. Access: 10 aug. 2022.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Introdução. *In*: BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação da informação**: conceitos e tecnologia das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013. Cap. 1, p. 1-14.

BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. Modelagem. *In*: BAEZA-YATES, Ricardo; RIBEIRO-NETO, Berthier. **Recuperação da informação**: conceitos e tecnologia das máquinas de busca. 2. ed. Porto Alegre: Bookman, 2013. cap. 2, p. 1-20.

BIBLIOTECA DE TESES E DISSERTAÇÕES. **Histórico**. Brasília: IBICT, [2022]. Available at: http://bdtd.IBICT.br/vufind/Content/history. Access: 26 aug. 2022.

BIBLIOTECA DE TESES E DISSERTAÇÕES. **Histórico**. Brasília: IBICT, [2022]. Available at: http://bdtd.IBICT.br/vufind/Content/history. Access: 26 aug. 2022.

CAFÉ, Luísa Chaves; MUNÕZ, Ivette Kafure; LEITE, Fernando César Lima. Usabilidade na recuperação da informação em acesso aberto: estudo da interação de usuários da pós-graduação com o Repositório Institucional da Universidade de Brasília. **Pesq. Bras. em Ci. da Inf. e Bib.**, João Pessoa, v. 10, n. 2, p. 032-046, 2015. Available at: https://www.brapci.inf.br/index.php/res/v/29771. Access: 28 jul. 2022.

CARDOSO, Olinda Nogueira Paes. Recuperação da informação. **INFOCOMP Revista de Ciência da Computação**, [Lavras], v. 2, n. 1, p. 33-38, 2004. Available at: https://infocomp.dcc.ufla.br/index.php/infocomp/article/view/46. Access: 14 jul. 2022.

GERHARDT, Tatiana Engel *et al*. Unidade 4 – estrutura do projeto de pesquisa. *In*: GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (org.). **Métodos de pesquisa**. Porto Alegre: Editora da UFRGS, 2009. p. 65-87. (Série Educação a Distância).

GOLDINHO, Natalia Bermudez. **Avaliação da recuperação da informação no repositório institucional da Universidade Federal do Rio Grande - FURG**: discussão sobre linguagem natural e controlada com ênfase na área de Biblioteconomia. 2014. 243 f. Trabalho de Conclusão de Curso (Graduação em Biblioteconomia) – Instituto de Ciências Humanas e da Informação - ICHI, Universidade Federal do Rio Grande, Rio Grande, 2014. Available at: http://repositorio.furg.br/handle/1/5941. Access: 26 aug. 2022.

IBICT. **TEDE/TEDE2**. Brasília, DF: Wiki IBICT, 2014. Available at: http://wiki.IBICT.br/index.php?title=TEDE/TEDE2&printable=yes. Access: 26 aug. 2022.

INOMATA, Danielly Oliveira; PINTRO, Sirlene. Agregar valor à serviços de unidades de informação. **Revista ACB**, Florianópolis, v. 18, n. 2, p. 1001-1017, sep. 2013. Available at: https://revistaacb.emnuvens.com.br/racb/article/view/913/pdf. Access: 22 jul. 2022.

LA REFERENCIA. **Desenvolvemos tecnologia para coletores nacionais**. [*S. l.*]: La Referencia, [2022]. Available at: https://www.lareferencia.info/pt/servicios/tecnologia. Access: 26 aug. 2022.

LEITE, Fernando *et al*. **Boas práticas para a construção de repositórios institucionais da produção científica**. Brasília: IBICT, 2012. Available at: http://livroaberto.IBICT.br/handle/1/703. Access: 26 aug. 2022.

LOURENÇO, Cíntia de Azevedo. **Análise do padrão brasileiro de metadados de teses e dissertações segundo o modelo entidade-relacionamento**. 2005. 164 f. Tese (Doutorado) –

| **22**

Universidade Federal de Minas Gerais, Escola de Ciência da Informação, Belo Horizonte, 2005. Available at: https://repositorio.ufmg.br/handle/1843/EARM-6ZGNZC. Access: 26 aug. 2022.

MACHADO, Jorge. Dados abertos e ciência aberta. *In*: Albagli, Sarita; MACIEL, Maria Lucia; ABDO, Alexandre Hannud (org.). **Ciência aberta, questões abertas**. Brasília: IBICT; Rio de Janeiro: UNIRIO, 2015. Cap. 9, p. 201-227. Available at: https://9x0o.short.gy/questõesabertas. Access: 10 jun. 2023.

MARCONI, Marina de Andrade; LAKATOS, Eva Maria. **Metodologia científica**. 7. ed. rev. ampl. São Paulo: Atlas, 2014. 225 p.

PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico**: métodos e técnicas da pesquisa e do trabalho acadêmico. 2. ed. Novo Hamburgo: Feevale, 2013.

REPOSITÓRIO UFPB. **DSpace help**. [João Pessoa]: UFPB, [2022]. Available at: https://repositorio.ufpb.br/jspui/help/index.html. Access: 26 aug. 2022.

SAMPAIO, Débora Adriano; LOUREIRO, José Mauro Matheus. Informação e memória na perspectiva da teoria ator rede. **PontodeAcesso**, Salvador, v. 13, n. 1, p. 47-64, apr. 2019. Available at: https://brapci.inf.br/index.php/res/download/124206. Access: 19 jun. 2022.

SANTOS JUNIOR, Ernani Rufino dos. **Repositórios institucionais de acesso livre no Brasil**: estudos delfos. 2010. 177 f. Dissertação (Mestrado) – Universidade de Brasília, Faculdade de Ciência da Informação, Brasília, DF, 2010. Available at: https://periodicos.unb.br/index.php/RICI/article/view/1977. Access: 11 jun. 2023.

SARACEVIC, Tefko. **Ciência da informação:** origem, evolução e relações. Belo Horizonte, 1996, p. 41-62.

SOUZA, Renato Rocha. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspect. ciênc. inf.**, Belo Horizonte, v. 11 n. 2, p. 161-173, maio./ago. 2006. Available at: https://brapci.inf.br/index.php/res/v/33177. Access: 18 jul. 2022.

TEDE PUCRS. **Sobre a Biblioteca Digital Brasileira de Teses e Dissertações (BDTD)**. [Porto Alegre]: PUCRS, [2022]. Available at: https://tede2.pucrs.br/tede2/static/bdtd.jsp. Access: 26 aug. 2022.

TEDE UFC. **Sistema de Publicação Eletrônica de Teses e Dissertações – TEDE**. [Fortaleza]: UFC, [2022]. Available at: http://www.teses.ufc.br/tde_ajuda/pt-BR/tde_index/index.php. Access: 26 aug. 2022.

TEDEBC UFMA. **Ajuda**. [São Luis]: UFMA, [2022]. Available at: https://tedebc.ufma.br/jspui/help/index.html. Access: 26 aug. 2022.

UNESCO. **Ciência aberta no Brasil**. Brasília, DF: UNESCO, 2023. Available at: https://www.unesco.org/pt/fieldoffice/brasilia/expertise/open-science-brazil. Access: 9 jun. 2023.

| **23**

WEITZEL, Simone da Rocha. Iniciativa de arquivos abertos como nova forma de comunicação científica. *In*: SEMINÁRIO INTERNACIONAL LATINO-AMERICANO DE PESQUISA EM COMUNICAÇÃO, 3, 2005, São Pulo. **Anais** [...]. São Paulo: ALAIC, 2005. Available at: http://eprints.rclis.org/6492/. Access: 22 jul. 2022.

WEITZEL, Simone da Rocha. O mapeamento dos repositórios institucionais brasileiros: perfil e desafios. **Encontros Bibli**, Florianópolis, v. 24, n. 54, p. 105-123, jan./apr. 2019. Available at: https://9x0o.short.gy/repositorios. Access: 11 jun. 2023.

| 24

APENDIX A. HEIs in the northeast region and their institutional repositories.

ID	HEI	IR
1	FEDERAL UNIVERSITY OF GRANDE DOURADOS (UFGD) FOUNDATION	UFGD Repository https://repositorio.ufgd.edu.br
2	FEDERAL UNIVERSITY OF HEALTH SCIENCES OF PORTO ALEGRE (UFCSPA) FOUNDATION	Non-BDTD participant
3	FEDERAL UNIVERSITY OF RONDÔNIA (UNIR) FOUNDATION	UNIR Institutional Repository https://ri.unir.br
4	FEDERAL UNIVERSITY OF ABC (UFABC) FOUNDATION	UFABC Institutional Repository
5	FEDERAL UNIVERSITY OF PAMPA (UNIPAMPA) FOUNDATION	Unipampa Institutional Repository https://dspace.unipampa.edu.br
6	FEDERAL UNIVERSITY OF TOCANTINS (UFT) FOUNDATION	UFT Institutional Repository http://repositorio.uft.edu.br/
7	FEDERAL UNIVERSITY OF THE SÃO FRANCISCO VALLEY (UNIVASF) FOUNDATION	Non-BDTD participant
8	UNIVERSITY OF INTERNATIONAL INTEGRATION OF AFRO-BRAZILIAN LUSOPHONY (UNILAB)	UNILAB Institutional Repository http://repositorio.unilab.edu.br
9	UNIVERSITY OF BRASILIA (UNB)	UNB Institutional Repository https://repositorio.unb.br/
10	FEDERAL UNIVERSITY OF BAHIA (UFBA)	UFBA Institutional Repository https://repositorio.ufba.br/
11	FEDERAL UNIVERSITY OF FRONTEIRA SUL (UFFS)	Institutional Repository https://rd.uffs.edu.br/
12	FEDERAL UNIVERSITY OF LATIN AMERICAN INTEGRATION (UNILA)	UNILA Institutional Repository https://dspace.unila.edu.br/
13	FEDERAL UNIVERSITY OF PARAÍBA (UFPB)	UFPB Institutional Repository https://repositorio.ufpb.br/
14	FEDERAL UNIVERSITY OF ALAGOAS (UFAL)	UFAL Institutional Repository https://www.repositorio.ufal.br/
15	FEDERAL UNIVERSITY OF ALFENAS (UNIFAL-MG)	Non-BDTD participant
16	FEDERAL UNIVERSITY OF CAMPINA GRANDE (UFCG)	Digital Library of Theses and Dissertations of UFCG http://bdtd.ufcg.edu.br/
17	FEDERAL UNIVERSITY OF CATALÃO (UFCAT)	Non-BDTD participant
18	FEDERAL UNIVERSITY OF GOIÁS (UFG)	Digital Library of Theses and Dissertations of UFG https://repositorio.bc.ufg.br/tede/
19	FEDERAL UNIVERSITY OF ITAJUBÁ (UNIFEI)	UNIFEI Institutional Repository https://repositorio.unifei.edu.br/jspui/
20	FEDERAL UNIVERSITY OF JATAÍ (UFJ)	Non-BDTD participant
21	FEDERAL UNIVERSITY OF JUIZ DE FORA (UFJF)	UFJF Institutional Repository https://repositorio.ufjf.br/jspui/
22	FEDERAL UNIVERSITY OF LAVRAS (UFLA)	UFLA Institutional Repository http://repositorio.ufla.br/jspui/
23	FEDERAL UNIVERSITY OF MATO GROSSO (UFMT)	UFMT Institutional Repository https://ri.ufmt.br/
24	FEDERAL UNIVERSITY OF MATO GROSSO DO SUL (UFMS)	UFMS Institutional Repository https://repositorio.ufms.br/
25	FEDERAL UNIVERSITY OF MINAS GERAIS (UFMG)	UFMG Institutional Repository https://repositorio.ufmg.br/
26	FEDERAL UNIVERSITY OF OURO PRETO (UFOP)	UFOP Institutional Repository http://www.repositorio.ufop.br/
27	FEDERAL UNIVERSITY OF PELOTAS (UFPEL)	UFPEL Institutional Repository http://guaiaca.ufpel.edu.br
28	FEDERAL UNIVERSITY OF PERNAMBUCO (UFPE)	UFPE Institutional Repository https://repositorio.ufpe.br/

29	FEDERAL UNIVERSITY OF RONDONÓPOLIS (UFR)	Non-BDTD participant
30	FEDERAL UNIVERSITY OF RORAIMA (UFRR)	Digital Library of Theses and Dissertations of UFRR http://www.bdtd.ufrr.br/
31	FEDERAL UNIVERSITY OF SANTA CATARINA (UFSC)	UFSC Institutional Repositoryhttps://repositorio.ufsc.br/xmlui/
32	FEDERAL UNIVERSITY OF SANTA MARIA (UFSM)	UFSM Digital Repository https://repositorio.ufsm.br/
33	FEDERAL UNIVERSITY OF SÃO CARLOS (UFSCAR)	UFSCAR Institutional Repository https://repositorio.ufscar.br/
34	FEDERAL UNIVERSITY OF SÃO JOÃO DEL REI (UFSJ)	Non-BDTD participant
35	FEDERAL UNIVERSITY OF SÃO PAULO (UNIFESP)	UNIFESP Institutional Repository https://repositorio.unifesp.br/
36	FEDERAL UNIVERSITY OF SERGIPE (UFS)	UFS Institutional Repository https://ri.ufs.br/
37	FEDERAL UNIVERSITY OF UBERLÂNDIA (UFU)	UFU Institutional Repository https://repositorio.ufu.br/
38	FEDERAL UNIVERSITY OF VIÇOSA (UFV)	UFV Institutional Repository https://www.locus.ufv.br/
39	FEDERAL UNIVERSITY OF ACRE (UFAC)	Non-BDTD participant
40	FEDERAL UNIVERSITY DO OF AGRESTE OF PERNAMBUCO (UFAPE)	Non-BDTD participant
41	FEDERAL UNIVERSITY OF AMAPÁ (UNIFAP)	Non-BDTD participant
42	FEDERAL UNIVERSITY OF AMAZONAS (UFAM)	Digital Library of Theses and Dissertations of UFAM https://tede.ufam.edu.br/
43	FEDERAL UNIVERSITY OF CARIRI (UFCA)	Non-BDTD participant
44	FEDERAL UNIVERSITY OF CEARÁ (UFC)	UFC Institutional Repository https://repositorio.ufc.br/
45	FEDERAL UNIVERSITY OF DELTA OF PARNAIBA (UFDPAR)	Non-BDTD participant
46	FEDERAL UNIVERSITY OF ESPÍRITO SANTO (UFES)	UFES Institutional Repository http://repositorio.ufes.br/
47	FEDERAL UNIVERSITY OF THE STATE OF RIO DE JANEIRO (UNIRIO)	Non-BDTD participant
48	FEDERAL UNIVERSITY OF MARANHÃO (UFMA)	Digital Library of Theses and Dissertations of UFMA https://tedebc.ufma.br
49	FEDERAL UNIVERSITY OF WESTERN BAHIA (UFBA)	Non-BDTD participant
50	FEDERAL UNIVERSITY OF WESTERN PARÁ (UFOPA)	UFOPA Institutional Repository https://repositorio.ufopa.edu.br/jspui/
51	FEDERAL UNIVERSITY OF PARÁ (UFPA)	UFPA Institutional Repository http://repositorio.ufpa.br/jspui/
52	FEDERAL UNIVERSITY OF PARANÁ (UFPR)	Digital Library of Theses and Dissertations of UFPR https://acervodigital.ufpr.br/
53	FEDERAL UNIVERSITY OF PIAUÍ (UFPI)	Non-BDTD participant
54	FEDERAL UNIVERSITY OF THE RECONCAVO OF BAHIA (UFRB)	UFRB Institutional Repository http://www.repositorio.ufrb.edu.br
55	FEDERAL UNIVERSITY OF RIO DE JANEIRO (UFRJ)	UFRJ Institutional Repository https://pantheon.ufrj.br/
56	FEDERAL UNIVERSITY OF RIO GRANDE (FURG)	FURG Institutional Repository http://repositorio.furg.br/
57	FEDERAL UNIVERSITY OF RIO GRANDE DO NORTE (UFRN)	UFRN Institutional Repository https://repositorio.ufrn.br/jspui/
58	FEDERAL UNIVERSITY OF RIO GRANDE DO SUL (UFRGS)	UFRGS Digital Repository https://lume.ufrgs.br/
59	FEDERAL UNIVERSITY OF SOUTHERN BAHIA (UFSB)	Non-BDTD participant

| 26

60	FEDERAL UNIVERSITY OF SOUTHERN AND SOUTHEASTERN PARÁ (UNIFESSPA)	Non-BDTD participant
61	FEDERAL UNIVERSITY OF VALES DO JEQUITINHONHA E MUCURI (UFVJM)	UFVJM Institutional Repository http://acervo.ufvjm.edu.br/jspui/
62	FEDERAL UNIVERSITY OF MINEIRO TRIANGLE (UFTM)	Digital Library of Theses and Dissertations of UFTM http://bdtd.uftm.edu.br/
63	FLUMINENSE FEDERAL UNIVERSITY (UFF)	UFF Institutional Repository https://app.uff.br/riuff/
64	FEDERAL RURAL UNIVERSITY OF AMAZON (UFRA)	UFRA Institutional Repository http://repositorio.ufra.edu.br
65	FEDERAL RURAL UNIVERSITY OF PERNAMBUCO (UFRPE)	Digital Library of Theses and Dissertations of UFRPE http://www.tede2.ufrpe.br/
66	FEDERAL RURAL UNIVERSITY OF RIO DE JANEIRO (UFRRJ)	Digital Library of Theses and Dissertations of UFRRJ https://tede.ufrrj.br/jspui/
67	FEDERAL RURAL UNIVERSITY OF SEMI-ARID REGION (UFERSA)	UFS Institutional Repository https://ri.ufs.br/
68	FEDERAL UNIVERSITY OF TECHNOLOGY OF PARANÁ (UTFPR)	UTFPR Institutional Repository http://repositorio.utfpr.edu.br/jspui/

Fonte: Dados da pesquisa (2022).

| 27