



TEXTO DIGITAL

Revista de Literatura, Linguística, Educação e Artes

Digital collections: practices of preservation and treatment of works

Acervos digitalizados: práticas de preservação e tratamento de obras

Ana Paula Nunes de Sousa^a; Emanuel Cesar Pires de Assis^b

^a Universidade Estadual do Maranhão, São Luís, Maranhão, Brasil - anapaulacxs123@outlook.com

^b Universidade Estadual do Maranhão, São Luís, Maranhão, Brasil - emanoel.uema@gmail.com

Keywords:

Literary collections.
Digital medium.
Search.
Preservation.

Abstract: Given the current context, an age of technological innovations, thinking about old books in the printed format and not correlating them to the new moment is perhaps not the best of the alternatives, especially if we consider paper's fragility. However, a question is necessary: What, after all, are the advantages of using digital support in the process of preserving a literary collection? Regarding the possible answers, we can say that there are many benefits gained from this process. Among them, we have the ease and speed of access to data and information, in addition to the democratization of knowledge. Thus, thinking of such advantages, the present study aims to present the possibilities of research in digital collections, as well as the preservation practices executed with literary collections in the city of Caxias-MA. In this article, we detail quantitative and qualitative researches that consist of data organization and systematization processes, available on the website Digital Library of Literature in Maranhão (<https://www.literaturamaranhense.ufsc.br>).

Palavras-chave:

Acervos Literários.
Suporte Digital.
Pesquisa.
Preservação.

Resumo: Diante do atual contexto, era de inovações tecnológicas, pensar o livro no formato impresso e não o correlacionar ao novo momento talvez não seja a melhor das alternativas, sobretudo se considerarmos sua fragilidade. No entanto, uma pergunta se faz necessária: Quais, afinal, as vantagens da utilização do suporte digital no processo de preservação de um acervo literário? Sobre as possíveis repostas, podemos dizer que muitos são os benefícios adquiridos com esse processo. Dentre eles, temos a facilidade e a velocidade de acesso a dados e informações, além da democratização do conhecimento. Assim, pensando em tais vantagens, o presente estudo tem como objetivo apresentar as possibilidades de pesquisas em acervos digitalizados, bem como as práticas de preservação realizadas junto a acervos literários da cidade de Caxias-MA. Uma pesquisa quantitativa e qualitativa que consiste em um processo organizacional e de sistematização de dados, disponibilizados em um sítio na internet, "O Portal Maranhão" (https://www.literaturamaranhense.ufsc.br/?locale=pt_BR).



Introduction

Many are the literary collections that today are vulnerable to the corrosive and destructive processes caused by agents of deterioration, which makes us think and question about the preservation strategies and what are the appropriate solutions to keep archives available and, at the same time, what can be done to safeguard the historical and cultural heritage they help to keep. Thinking about the economic and social development of a given community and not establishing actions that emphasizes material, cultural and historiographic preservation of their memories is, in short, an attempt to erase identity and it is also an intellectual imprisonment.

Thus, one of the possible actions to safeguard literary collections is the process of digitalization, which is an important tool for the preservation and dissemination of the literary memory of a community. In addition to facilitating access to information, the systematization and digitalization of data benefit its longevity, allowing access to the content without the need to handle the original document, making it available for remote consultation and at the reach of other researchers. Also, the recognition and appreciation of authors, if we consider the barriers and weaknesses that revolve around them and their literary creations, destined, most of the time, to obliviousness and abandonment, is reinforced, bringing up writers and works that generally are not within the scope of some researches.

With the objective of safeguarding, preserving and making available the data of literary production in Maranhão, the *Núcleo de Pesquisa em Literatura, Arte e Mídia – LAMID – UEMA*¹, in partnership with the *Núcleo de Pesquisas em Informática, Literatura e Linguística – NUPILL*², from the Federal University of Santa Catarina - UFSC, executes a detailed work with the collections of the *Academia Caxiense de Letras*³, of the *Instituto Histórico e Geográfico de Caxias – IHGC*⁴, and the Benedito Leite Library. Such work

¹ Literature, Art and Media Research Center – LAMID.

² Informatics, Literature and Linguistics Research Center – NUPILL.

³ Academy of Letters of the city of Caxias – Maranhão.

⁴ Historical and Geographical Institute of Caxias – IHGC.

is funded since 2015 by *Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Maranhão – FAPEMA*⁵.

Therefore, thinking about the advantages acquired with this initiative, in which the book is understood from a new support, the digital one, we will highlight, in this study, the possibilities of research in digital collections, as well as the practices of preservation and dissemination of literary memory from Maranhão implemented with the aforementioned collections. Such organizational and data systematization process has been made available on a website: the Portal Maranhão, (<https://www.literaturamaranhense.ufsc.br>).

Methods

The methodological procedures performed are based on two basic strategies. The first is to focus on organizing and sharing the information gathered in the collection, divided into four succinct stages: the first consists of the process of cataloging and dividing the works, which are initially separated and classified according to genre and conservation condition. Soon after, the classified works are systematized in Biblivre (<http://biblivre.org.br/index.php/en/>), a computer program that serves for the organization and systematization of data from a collection. The second stage consists of scanning works and documents. After selected and organized, the documents are submitted to the digitization process. In this procedure, we use a planetary scanner. The third stage is what we call the correction and updating of the spelling of digitized works. After the digitalization, the works go through the conversion process, made possible by the ABBYY software (<https://www.abbyy.com/en-apac/>), in an effort to update their original format (.pdf) to a correctable format (.docx). When the mistakes generated by the OCR program are removed, the works are finally updated. Finally, the fourth stage, still of the first strategy, consists of the process of revising the material and feeding the Portal Maranhão database. In this stage, all the organized data are made available; the works that went through the process of conversion and updating of the spelling are made available in HTML format. The second strategy is understood as the research itself, when the particularities of the life and works of the writers from Maranhão are observed.

⁵ Foundation for the Support of Research and Scientific and Technological Development of Maranhão – FAPEMA.

Results

Founded on August 15, 1997, the Academia Caxiense de Letras, or Coelho Neto's House, aggregates a diverse number of poets and prosaists and today is one of the most important memory institutions in the State of Maranhão, having a collection formed by more than three thousand works. The collection is divided and organized in works by national and international authors, in the areas of Literature, History, Education, Law, Sociology, Philosophy, Psychology and also the newspapers *Folha de Caxias* (1965-1973), *O Pioneiro* (1990-1994), *Jornal da Cidade* (1997-2001), among others.

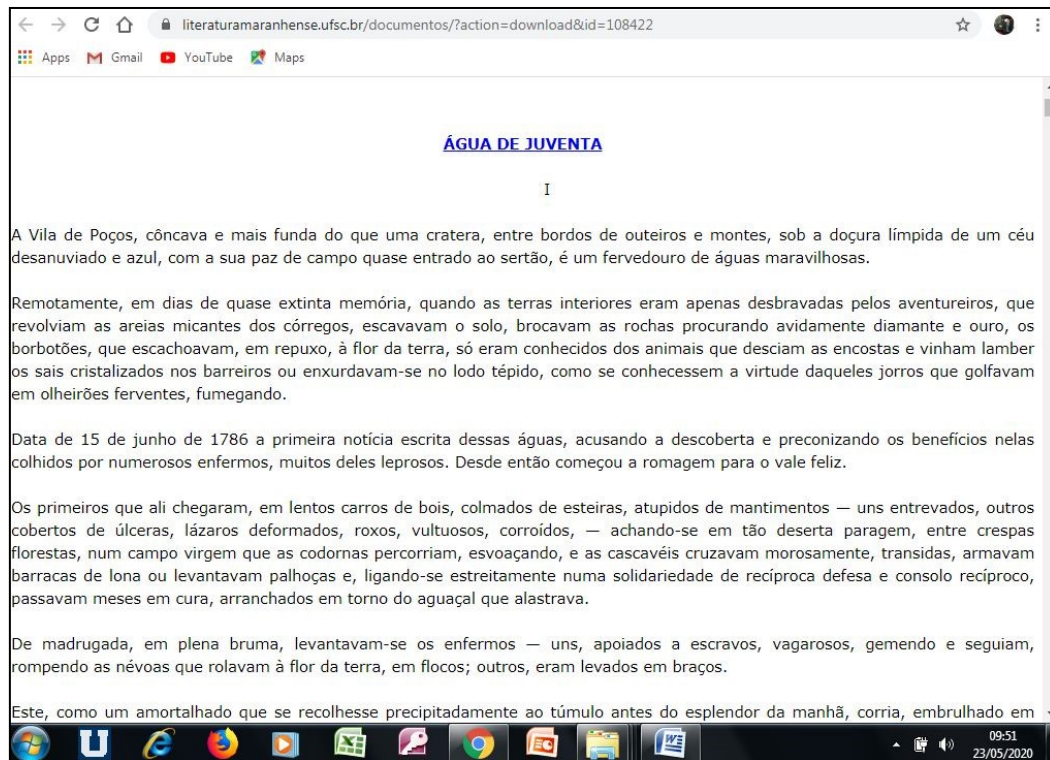
The work of preserving and disseminating the data present in this collection, as stated above, follows two basic strategies. The first process is based on the selection, cataloging and systematization of the collected and organized information, and as a previous result, we can say that 95% of the available documents are already registered, including books, magazines, and newspapers.

As examples, we can mention the works of Coelho Neto (1864-1934) and Humberto de Campos (1886-1934), two writers that were born in Maranhão state and that were very famous when alive, but now are almost forgotten in the literary studies. Regarding Coelho Neto's work available on Portal Maranhão, 114 documents (novels, books of short tales, political speeches, plays and essays) are registered and almost half of it can be read online or downloaded. *Rapsódias* (1891), *Turbilhão* (1906), *Miragem* (1895), *Romanceiro* (1898), *a Conquista* (1899), and *Banzo* (1927) are some of the works that can be found in HTML and PDF formats.

About Humberto de Campos' work, *Da Seara de Booz* (1918), *Notas de um diarista* (1935), *Memórias Inacabadas* (1935), *Últimas Crônicas* (1936) and *Fragmentos de um diário* (1962) are some works that can be found, from a total of 55.

Below is an example of a work converted into HTML format, *Água de Juventa*, published in 1905 by Coelho Neto, with the spelling updated.

Figure 1 - Work in html format.



Source: (<https://www.literaturamaranhense.ufsc.br/>).

The proposal for the creation of the Portal Maranhão arose from a project developed at *NUPELL* and at *LAPESD - Distributed Systems Research Laboratory*. The project, named Digital Library of Portuguese-Language Literatures, main objective is to create an environment for reading and teaching-learning literature. The Portal Maranhão currently has 3,770 works and 878 registered authors, as well as 534 digitized files and 4,852 registered publishers and journals⁶.

Therefore, given the efforts presented in the initial stages of the preservation work of the literary collection of the State of Maranhão, we can now discuss, even if in a summarized way, the second strategy. In this strategy we highlight the results of a research that has been accomplished with the works of Humberto de Campos, one of the most illustrious writers of Brazilian literature of the 20th century, especially of the 1930s. Author of a diverse number of literary works, Humberto de Campos is considered by his

⁶ These are the numbers available in May 19, 2020.

contemporary critics as one of the best chroniclers of his time, alongside important writers like Coelho Neto.

According to a search made at the Digital Library of Portuguese-Language Literatures (<https://www.literaturabrasileira.ufsc.br>) and on the website of the Brazilian Academy of Letters (<http://www.academia.org.br/memoria-da-abl/apresentacao>), Humberto de Campos Veras, son of Joaquim Gomes de Faria Veras and Ana de Campos Veras, was born on October 25, 1886, in the city of Miritiba. He was a journalist, politician, critic, essayist, short story writer, chronicler, memorialist and poet. He was also a participative and collaborating member of different periodicals, such as *Folha do Norte*, *Província do Pará* and *O Imparcial*, in addition to the newspapers *Gazeta de Notícias* and *Correio da Manhã*. However, very early on, he was victim of a parasitic disease known as elephantiasis, dying on December 5, 1934, in the city of Rio de Janeiro.

Regarding the style of the Miritiba-born writer, Scheibe (2008) highlights that the author's inspirations were ancient writers. His literary and bibliographic references were based on the ideas of Pericles of Athens, Horace of Rome, Ferdowsi of Persia and Arab legends. Almost all of his writings had classic characteristics, with texts that overlap and blend with the magical world of poetry and aspects of reality.

The work with Campos' texts, at the moment, comes from an attempt to prove the authenticity of the work *Crônicas de Além-Túmulo*, published in 1937, the first of a set of 12 attributed to the writer from Maranhão by the most famous Brazilian medium, Francisco Cândido Xavier. The theoretical and critical line chosen to operate in the research are stylistic and stylometric studies, using a computerized tool, the Hyperbase software, capable of attaining, as stated by Cúrcio (2006), word frequency, lexical distance, vocabulary growth, lexical richness, and various other textual features, made possible through fast computer manipulation.

An example of a research performed using this tool is the monographic work by Daniel Lopes (2017), entitled *A caracterização do estilo de escrita de Coelho Neto no romance Miragem (1895), a partir de uma abordagem estilométrica*. Where the researcher presents important concepts about stylistics and stylometry, highlighting ideas from authors such as José Lemos Monteiro (2009), Pierre Guiraud (1970) and Nilce Sant'anna

Martins (2000), scholars who draw a fine line in the emergence of the stylistic studies focused on the study of language, whose object of study is style.

In Monteiro's (2009) perspective, stylistics acquired the status of a discipline, entering an effervescent phase, in an effort to assert itself before other areas that disputed the same object: style. Two directions divide its trajectory. The first was concentrated more on the components of the speech, defined as descriptive stylistics, idealized by Charles Bally, and the second, inclined towards intuition, being labeled by genetic or idealistic stylistics, linked to the linguist Leo Spitzer.

Regarding its possible developments, Lopes (2017) states that stylistics has acquired, over time, some particularities, subdividing into structural, generative, rhetorical, poetic, semiotic and statistics, resulting in the ideas of stylometry. For him, literary stylometry, also seen as an offshoot of Bally's stylistics, constitutes itself as a resource, a methodology that helps, in a more precise way, the researcher in the search for stylistic traits.

Freitas (2007) states that, in Brazil, there has been for some time a scientific community in the areas of Literature and Linguistics discussing the role of new technologies and the changes brought about by it in the scope of teaching literature, in the dissemination of literary texts, the use the world wide web, as well as its role in literary production made in digital media and designed to be consumed in this type of support.

According to Lopes (2017), the computerized tools come to assist the researchers, helping them to accelerate the investigations in the field of literature. That is, in the field of statistical studies of literary texts, the researcher has the function and task of analyzing and interpreting the data that computerized programs make available. The scholar states that the achievements and data interpretation solutions that such programs offer, without escaping criticism, make the stylist researcher's work faster and more viable.

Authorship attribution: initial results

For an appreciation of our approach, we now turn to the initial results of our research, we will present, for a better understanding, the corresponding image of our corpus:

Figure 2 - Corpus Base.



Source: *Hyperbase*.

The corpus consists of 9 texts, distributed between Humberto de Campos, Chico Xavier and contrasted with four texts by *Machado de Assis* - *Últimas Crônicas* (CAMPOS, 1962), *Da Seara de Booz* (CAMPOS, 1918); *The Messengers* (XAVIER, 1944); *Crônicas de Além-Túmulo* (XAVIER, 1937); *Lição* (XAVIER, 1966); *Crônicas* (ASSIS, 1862), *Balas de Estalo* (ASSIS, 1883), *A Semana* (ASSIS, 1892) and *Paixão* (ASSIS, 1904).

For this, as a brief understanding, the code is presented in the corpus as the first letter of the name of the work and the name of the author; in the case of the analyzed text, *Crônicas de Além-Túmulo*, we work with the initial hypothesis of being authored by Chico Xavier. It is also worth noting that, all the texts selected belong to the prose category, the majority of which are chronic, inserted and organized from the following nomenclatures: **UltC**; **Seara**; **Além**; **Mens**; **Lição**; **Crôn**; **Balas**; **Semana**; **Paixão**.

The corpus, as a whole, has 452,759 occurrences⁷. The image below shows the layout of the texts, the number of occurrences, words⁸ and its extension.

Figure 3 - Occurrences, words and corpus extension.

Extensão e Prob	Nº	TÍTULO	OCORRÊNCIAS	VOCÁBULOS	Prob P	Prob Q	ABREVIADO	CÓDIGO
	1	ÚltC	40677	7825	.0898	.9102	ÚltC	UH
Riqueza e hapax	2	Seara	42784	9067	.0945	.9055	Seara	SH
	3	Além	47647	8972	.1052	.8948	Além	AC
Cresc. cronol.	4	Mens	71379	9922	.1577	.8423	Mens	MC
	5	Lição	35346	7870	.0781	.9219	Lição	LC
Cresc. inverso	6	Crôm	22905	4706	.0506	.9494	Crôm	CM
	7	Balas	40285	6849	.089	.911	Balas	BM
	8	Semana	150002	15831	.3813	.6687	Semana	SM
	9	Paixão	1725	659	.0038	.9962	Paixão	PM
		TOTAL	452750	35071				

Source: *Hyperbase*.

From the data presented above, we can say that Humberto de Campos' vocabulary extension in **ÚltC** (1962) is 7,825; the text **Seara** (1918), also by the writer from Maranhão, has 9,067 words. The texts **Além** (1937), **Mens** (1944), both from the medium Chico Xavier, have 8,972 and 9,922 words, respectively.

Thus, in terms of extension of words, the texts of the analyzed authors, Humberto de Campos and Chico Xavier do not distance from each other. Considering the number of occurrences **Mens** (1944) is the most extensive one, with a total of 71,379 occurrences, which explains its extension of the vocabulary. While the others, **ÚltC** (1962), **Seara** (1918) and **Além** (1937), present 40,667, 42,784 and 47,647, respectively.

⁷ Total number of words or forms in the base corpus.

⁸ Number of words or forms that appear only once in the base corpus.

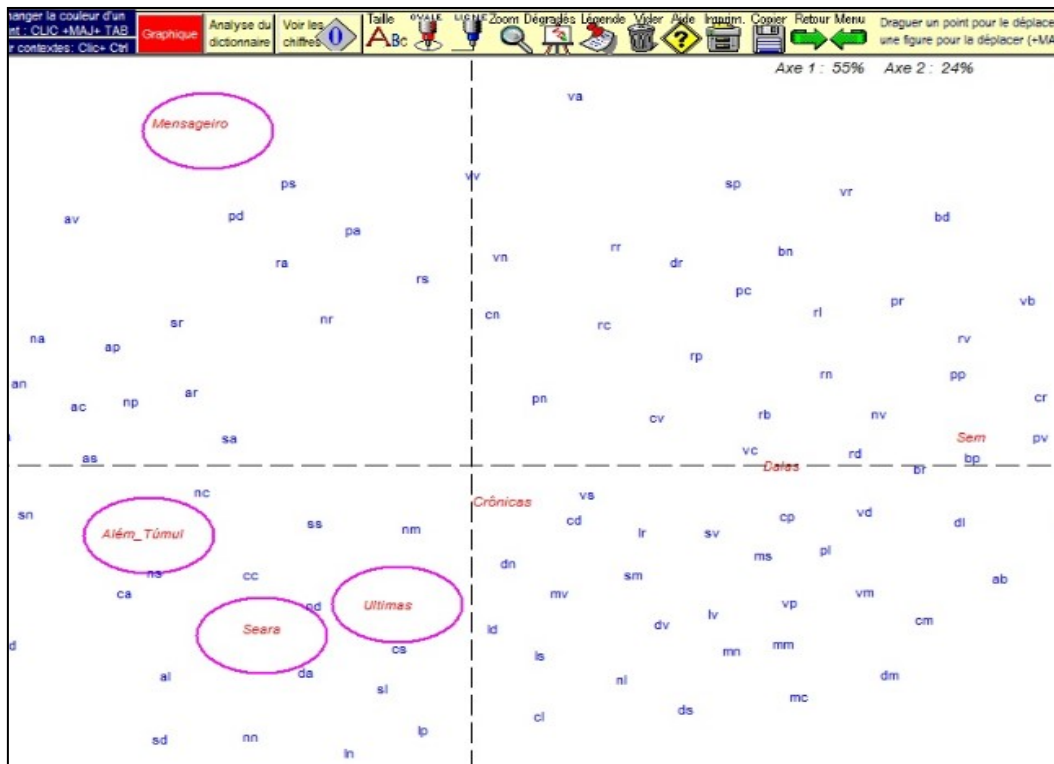
For a better visibility of the highlighted data, below is a chart that corresponds to the distribution of the number of occurrences and the number of words in each text.

Chart 1 - Number of occurrences and number of words.

Code	Occurrences (N)	Words (V)
UltC	40.667	7.825
Seara	42.784	9.067
Além	47.647	8.972
Mens	71.379	9.922
Lição	35.346	7.870
Crôn	22.905	4.706
Balas	40.285	6.849
Semana	150.002	15.831
Paixão	17.25	659
TOTAL	452.750	35.071

Source: Hyperbase.

A first interpretation can be made of one of the excerpts of the research. The factorial analysis of the bicodes consists of mapping the grammatical combinations, taking in consideration simultaneous use of two grammatical codes. In such analysis, in terms of style, it is verified an approximation between the analyzed text, *Crônicas de Além-Túmulo*, attributed by Chico Xavier to the author Humberto de Campos, with those of proven authorship (*Da Seara de Booz* and *Últimas Crônicas*), as shown in the image below:

Figure 4 - Factorial analysis of the bicodes.

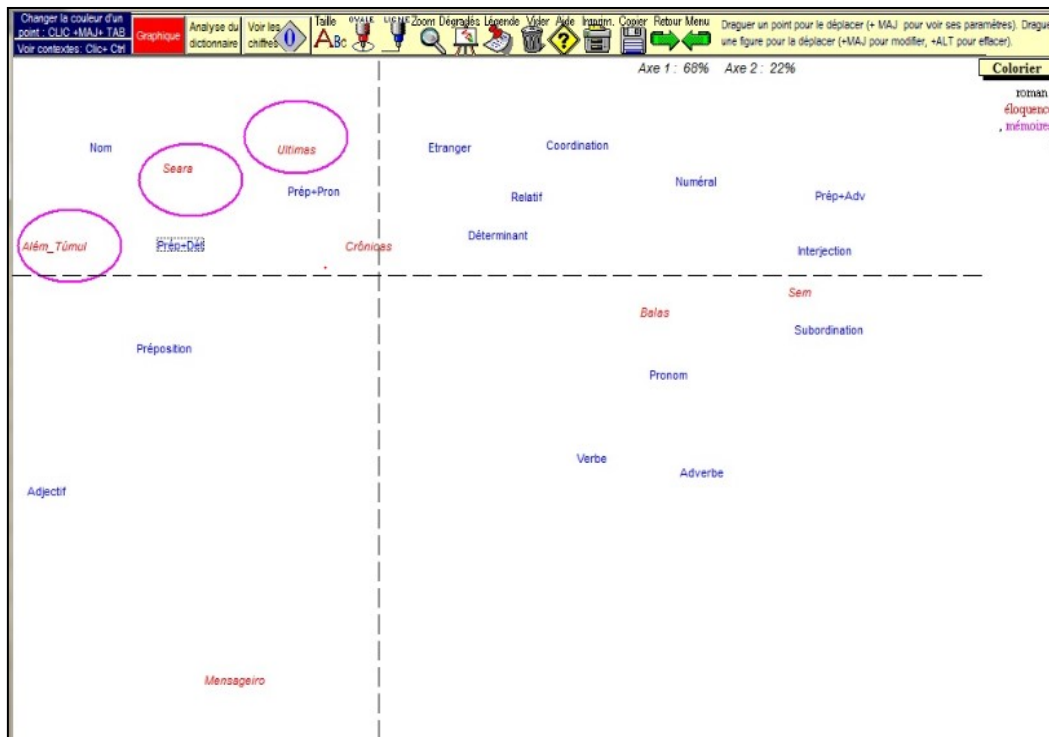
Source: *Hyperbase*.

The texts of the analyzed authors, arranged in the same quadrant, present similar grammatical distribution, in terms of the use of bicodes. The author or “authors” tends to use the following grammatical elements in the composition of sentences: preposition + noun; nouns + conjunction; noun + preposition; determinant + preposition; conjunction + adjective; determinant + adjective; noun + numeral, which for the study in question becomes an important information.

Another significant data for the study of authorship attribution is the grammatical distribution of the texts; in the case of Campos, the literary criticism accentuates the style of his writing as using a high rate of nouns (ROCHA, 2012).

The image below shows us once again the approximation of Campos' texts with the work *Crônicas de Além-Túmulo*, as well as the grammatical forms that most appear in *Da Seara de Booz* (1918) and *Últimas Crônicas* (1962). According to the data, the present study was able to identify that the text by Xavier exhibits more preposition + determinant; preposition + pronoun; and nouns, dialoguing with the stylistic study done by the critic (ROCHA, 2012) regarding the style of Campos, as previously highlighted.

Figure 5 - Grammatical analysis.



Source: *Hyperbase*.

Therefore, from the data presented, we can say that this is an ongoing research and that somehow it has perceived stylistic approximations and similarities between the analyzed text *Crônicas de Além-Túmulo*, psychographed by Chico Xavier, and the texts by Campos.

Besides authorship attribution, several other literary researches can be done using digital collections. Assis, Sousa and Silva (2020), Assis and Lopes (2019), Assis (2017) and Assis and Sousa (2017) are just some examples of how literary studies face a field with huge opportunities.

Discussion

Returning to the premise engaged at the beginning of our article on what are the advantages of using digital support in the process of preserving a literary collection, we can say that the changes regarding the perception of the printed book are notorious, as well as the significant advantages to preserve a literary collection from a digital perspective.

Santarém Segundo (2014), in the work *O uso de elementos semânticos no processo de recuperação da informação em ambientes digitais*⁹ says that it is currently possible to verify that several institutions and organizations have made decisions regarding processes that involve the use of information in digital format. The researcher also points out that most of these institutions have been concerned with creating documents that while persisting in the paper format can also be converted to digital format through digitization processes, considering that the book, in printed format, is materialized utilizing digital tools. Likewise, we see that,

Technology has been opening space for the process of digitizing documents, enabling consultation of the digital library. It is a new way to access content; it is a new format for the document, digital format, which greatly facilitates the search and the remote access to users (BRAGA e DIEMER, 2010, p. 10, our translation).

Therefore, given the current context of technological developments, the process of digitizing bibliographic collections appears as an alternative for preserving and facilitating access to information, in which the main line of defense works with the argument that it will benefit its longevity. The digital data could also allow access to content without the need to handle the original document, in addition to facilitating contact with the books, making them available for remote consultation and at the reach of other researchers, as highlighted by Greenhalgh (2011).

Valle e Araújo (2005) point out that in the universe of conventional conservation techniques, preservation and access are dimensions that are not only distinguishable but also often opposed. Frequently, the only way to guarantee the preservation of an item is to reduce its circulation. However, with the application of digital technology, this scenario is radically transformed, as these dimensions become related and cooperative.

According to the researchers, digitization brings countless possibilities to the universe of collections preservation. However, its application in artifacts of permanent value must be conducted with care, accompanied by a long-term strategy, under the consequence of placing the collection at the mercy of the fragility of digital technology. According to

⁹ The use of semantic elements in the information retrieval process in digital environments

them, the issue of digital longevity and access must be addressed in any reformatting plan for the digital media.

In another way, Reifschneider (2008), also revolving the problem of the "Rare Works" objects, raises the following question: how to promote access to rare works and preserve them so access can be continued? Santarém Segundo (2014) presents a possible solution for the documents and rare works promotion and preservation. The researcher recommends the digital repositories, information environments constituted from and by technological tools.

Digital repositories are information environments created of technology tools/platforms, which can also be called information systems, capable of receiving deposits of digital objects, in various formats, whether made by the authors themselves or by teams trained for this purpose, in order to store and organize these objects so that it can be recovered and mainly preserved for a long time (SANTARÉM SEGUNDO, 2014, p. 112-113, our translation).

Regarding the basic concepts of storage and preservation, Boeres and Farias (2012) say that these remain applicable in the new professional culture, with computational tools replacing the pencil and paper, generating new routines that allow other processes of storage and recovery. This concept becomes even more valid as the importance and fragility of printed literary collections are perceived. There are many possibilities for corrosion and destruction of this cultural heritage, exposed and subjected to the actions of deterioration agents, such as humidity, the temperature and the actions of man, mainly for the inadequate handling and lack of technical knowledge with the document.

Therefore, understanding all the changes and resignifications of the process of safeguarding and preserving printed literary collections, Sousa, Correia and Assis (2018), in *A Digitalização da Literatura Maranhense: O Portal Maranhão* present important ideas and strategies to guarantee the longevity of Maranhão's literary memory. According to the researchers, we live in a new era, an era of digital culture. And literature, in this way, is included. Today, we have new horizons and possibilities for research and literary studies.

Conclusion

As initially stated, the goal of this article is to attempt to highlight the possibilities of research in digital collections. We also emphasized the preservation practices implemented with collections, a work divided and strategically organized, understood as a selective process and systematization of data, in which information about life and work of authors from Maranhão is made available on Portal Maranhão website.

Regarding the work of safeguarding literary collections, Mota (2014) says that archival institutions of public scope, considered as instruments of democratization of information, have rich documents with social interest content in their collections that often are not accessed due to poor conservation conditions, reduced access, as well as the lack of knowledge of society about its existence.

Therefore, from such information and realizing the fragility of the printed literary collections, as well as the possibilities of corrosion and destruction of this cultural heritage, that are often exposed and subjected to the actions of agents of deterioration, we accentuate that the need and importance of the digitization process becomes evident. In addition to the dissemination and availability in digital support, there is also the material preservation of the original printed document, since there is no need for direct contact with it, which will prolong its longevity. With the availability of these works on the internet, there is what we call the democratization of knowledge, as the researcher can easily have quick and remote access to these documents, many considered rare and out of reach to most users.

REFERENCES

ASSIS, Emanuel César Pires de; CORREIA, Marcus Vinicius Sousa; SOUSA, Ana Paula Nunes de. A Digitalização da Literatura Maranhense: O Portal Maranhão. *Revista Letra Magna*, Ano 14, n.23, p. 803-815, 2º Semestre, 2018. Available in: http://www.letramagna.com/artigos_23/artigo_lit_03_23.pdf. Accessed in April 16, 2020.

ASSIS, Emanuel Pires de. Leitura de literatura em meio digital: a mediação pelo corpo. *Texto Digital*, v. 13, n. 1, p. 141-156, 2017. Available in:

<https://periodicos.ufsc.br/index.php/textodigital/article/view/18079288.2016v13n1p141>. Accessed in April 20, 2020.

ASSIS, Emanuel Cesar Pires de; SOUSA, Ana Paula Nunes de; SILVA, Alanna Costa da. Literatura digitalizada: preservação e divulgação da memória literária maranhense. *Confluências Culturais*, v. 9, n. 1, p. 181-191, 2020. Available in: <http://periodicos.univille.br/index.php/RCCult/article/view/813>. Accessed in April 20, 2020.

ASSIS, Emanuel Cesar Pires; LOPES, Daniel. A estatística textual computadorizada e a literatura brasileira: uma análise do romance "Miragem", de Coelho Neto. *Studia Iberytyczne*, v. 18, p. 259-270, 2019. Available in: <https://www.journals.akademicka.pl/si/article/view/1047>. Accessed in April 20, 2020.

ASSIS, Emanuel Cesar Pires; SOUSA, Andressa Silva. "Efeito de Real" Versus Sobrenatural: um conflito necessário à construção da fantasmaticidade em *Esfinge*, De Coelho Neto. *Revista De Letras-Juçara*, v. 1, n. 1, p. 144-161, Jul, 2017. Available in: <http://ppg.revistas.uema.br/index.php/jucara/article/view/1324>. Accessed in April 15, 2020.

BRAGA, Paula Dantas; DIEMER, Vanessa Maria Almeida. Digitalização de obras raras: estudo comparativo do Senado Federal e do Supremo Tribunal Federal. 2010. 90 f. Trabalho de Conclusão de Curso – Universidade de Brasília – Faculdade de Ciência da Informação, Brasília, 2010. Available in: https://bdm.unb.br/bitstream/10483/12111/1/2010_VanessaDiemer_PaulaBraga.pdf. Accessed in July 22, 2019.

BOERES, Sonia Araujo de Assis; FARIA, Ana Carolina Cintra. A preservação digital na biblioteca central da Universidade de Brasília. In: *Ibict*, Brasília, DF, v.41, p.175-183, jan./abril., 2012. Available in: <http://revista.ibict.br/ciinf/article/view/1363>. Accessed in July 23, 2019.

CÚRCIO, Verônica Ribas. Estudos estatísticos de textos literário. *Texto Digital*, Florianópolis, ano, n. 2, p. 1-15, dezembro, 2006. Available in: <https://periodicos.ufsc.br/index.php/textodigital/article/view/1382>. Accessed in June 16, 2019.

DAMASCENO, Priscilla Mota. Digitalização de acervos fotográficos e o desafio entre o real e o ideal: uma análise a partir das necessidades do memorial do TRT/PB. 2014. 59 f. Trabalho de conclusão de curso – Universidade Federal da Paraíba, João Pessoa, 2014. Available in: <https://repositorio.ufpb.br/jspui/handle/123456789/1187>. Accessed in Jan 28, 2020.

DI GIORGI, F. V. *Por que filosofar?* In: Cadernos PUC, n.1, p. 79-80. São Paulo: EDUC/ Cortez Editora, 1980.

FREITAS, Deise. A composição do estilo do contista Machado de Assis. 2009. 204 f. Tese (Doutorado em Literatura) – Programa de Pós-Graduação em Literatura, Universidade Federal de Santa Catarina, Florianópolis, 2009. Available in: <https://repositorio.ufsc.br/xmlui/handle/123456789/90720>. Accessed in: Jan 28, 2020.

GREENHALGH, Rafael Diego. Digitalização de obras raras: algumas considerações. *Perspectivas em Ciência da Informação*, v. 16, n. 3, p. 159-167, jul./set., 2011. Available in: <https://doi.org/10.1590/S1413-99362011000300010>. Accessed in Dec 17, 2019.

LOPES, Daniel. A caracterização do estilo de escrita de Coelho Neto no romance *Miragem* (1895), a partir de uma abordagem estilométrica. 2017. 115 f. Trabalho de Conclusão de Curso – Universidade Estadual do Maranhão, Caxias-MA, 2017.

REIFSCHNEIDER, Oto Dias Becker. A importância do acesso às obras raras. *Revista Ibero-americana de Ciência da Informação* (RICI), v. 1, n. 1, p.67-76, jan./jun.2008. Available in: <https://www.brapci.inf.br/index.php/res/v/70651r>. Accessed in Dec 17, 2020.

ROCHA, Alexandre Caroli. Complicações de uma estranha autoria (O que se comentou sobre textos que Chico Xavier atribuiu a Humberto de Campos). *Ipotesi*, Juiz de Fora, v.16, n.2, p. 25-36, jul./dez. 2012. Available in: <http://www.ufjf.br/revistaipotesi/files/2011/05/CAP02-25-36.pdf>. Accessed in July 15, 2019.

SEGUNDO, José Eduardo Santarem. O uso de elementos semânticos no processo de recuperação da informação em ambientes digitais. *Texto Digital*, Florianópolis, v. 13, n. 2, p. 93-111, jul./dez. 2017. Available in: <https://periodicos.ufsc.br/index.php/textodigital/article/view/18079288.2017v13n2p93>. Accessed in July 15, 2019.

SOUSA, Ana Paula Nunes de; ASSIS, Emanuel César Pires de. De Miritiba à manchete de jornais: Humberto de Campos e o caso mediúnico. In: *I congresso Nacional de Linguística Aplicada e IV Encontro Nacional de Ficção Discurso e memória*, São Luís: EDUFMA, p. 52-65, 2018.

SCHEIBE, Roberta. *A crônica e seus diferentes estilos na obra de Humberto de Campos*. Imperatriz-MA: Ética, 2008.

VALLE, Eduardo; Araujo, Arnaldo. Digitalização de acervos, desafios para o futuro. *Revista do arquivo público mineiro*, Belo Horizonte, v. 4, jul-dez. 2005. Available in:

https://www.siaap.cultura.mg.gov.br/acervo/rapm_pdf/Digitalizacao_de_acervos_desafio_para_o_futuro.PDF. Accessed in Mar 15, 2020.

NOTAS DE AUTORIA

Ana Paula Nunes de Sousa (anapaulacxs123@outlook.com) é acadêmica do curso de Graduação em Letras da Universidade Estadual do Maranhão - UEMA.

Emanoel Cesar Pires de Assis (emanoel.uema@gmail.com) é Doutor em Literatura (UFSC). Docente na Universidade Estadual do Maranhão - UEMA.

Como citar este artigo de acordo com as normas da revista?

SOUSA, Ana Paula Nunes de; ASSIS, Emanoel Cesar Pires de. Digital collections: practices of preservation and treatment of works. *Texto Digital*, Florianópolis, v. 16, n. 2, p. 43-60, 2020.

Contribuição de autoria

Ana Paula Nunes de Sousa: concepção e elaboração do manuscrito; análise de dados; discussão dos resultados; revisão e aprovação.

Emanoel Cesar Pires de Assis: concepção e elaboração do manuscrito; análise de dados; discussão dos resultados; revisão e aprovação.

Financiamento

Fundação de Amparo à Pesquisa e ao Desenvolvimento Científico e Tecnológico do Estado do Maranhão - FAPEMA

Consentimento de uso de imagem

Figure 1 - Work in html format. Source: (<https://www.literaturamaranhense.ufsc.br/>).

Figure 2 - Corpus Base. Source: Hyperbase.

Figure 3 - Occurrences, words and corpus extension. Source: Hyperbase.

Figure 4 - Factorial analysis of the bicodes. Source: Hyperbase.

Figure 5 - Grammatical analysis. Source: Hyperbase.

Aprovação de comitê de ética em pesquisa

Não se aplica.

Licença de uso

Este artigo está licenciado sob a Licença Creative Commons CC-BY. Com essa licença você pode compartilhar, adaptar, criar para qualquer fim, desde que atribua a autoria da obra.

Histórico

Recebido em: 25/05/2020.

Aprovado em: 08/10/2020.