UNIVERSIDADE FEDERAL DO RIO DE JANEIRO

DANIEL MAIA

HUMANIDADES DIGITAIS: uma Emergência Teórico-Metodológica na Representação e Descoberta de Conhecimento para a História das Ciências no Brasil

Rio De Janeiro

2021

Daniel Maia

HUMANIDADES DIGITAIS: uma Emergência Teórico-Metodológica na Representação e Descoberta de Conhecimento para a História das Ciências no Brasil

Tese de Doutorado apresentada ao Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia/HCTE, Universidade Federal do Rio de Janeiro, como requisito parcial para a obtenção do título de Doutor em História das Ciências e das Técnicas e Epistemologia.

Orientadora: Regina Maria Macedo Costa Dantas

Rio De Janeiro

2021

CIP - Catalogação na Publicação

MM217h

Maia, Daniel
HUMANIDADES DIGITAIS: uma Emergência Teórico Metodológica na Representação e Descoberta de Conhecimento para a História das Ciências no Brasil / Daniel Maia. -- Rio de Janeiro, 2021.
128 f.

Orientadora: Regina Maria Macedo Costa Dantas.
Tese (doutorado) - Universidade Federal do Rio de Janeiro, Decania do Centro de Ciências Matemáticas e da Natureza, Programa de Pós-Graduação em História das Ciências e das Técnicas e Epistemologia, 2021.

1. História das Ciências. 2. Transdisciplinaridade. 3. Humanidades Digitais. I. Maria Macedo Costa Dantas, Regina, orient. II. Título.

Elaborado pelo Sistema de Geração Automática da UFRJ com os dados fornecidos pelo(a) autor(a), sob a responsabilidade de Miguel Romeu Amorim Neto - CRB-7/6283.

DANIEL MAIA

HUMANIDADES DIGITAIS: UMA EMERGÊNCIA TEÓRICO-METODOLÓGICA NA REPRESENTAÇÃO E DESCOBERTA DE CONHECIMENTO PARA A HISTÓRIA DAS CIÊNCIAS NO BRASIL

Tese submetida ao corpo docente do Programa de História das Ciências e das Técnicas e Epistemologia (HCTE), da Universidade Federal do Rio de Janeiro (UFRJ), como parte dos requisitos necessários à obtenção do grau de Doutor em História das Ciências, das Técnicas e Epistemologia.

Aprovada em: 29 de setembro de 2021

Profª. Drª. Regina Maria Macedo Costa Dantas
Universidade Federal do Rio de Janeiro

Prof. Dr. Paulo Vinicius Aprígio da Silva
Colégio Pedro II

Profª. Drª. Maria José Veloso da Costa Santos
FACC/Universidade Federal do Rio de Janeiro

Prof. Dr. Eduardo Nazareth Paiva
Universidade Federal do Rio de Janeiro

Prof. Dr. Rundsthen Vasques de Nader
Universidade Federal do Rio de Janeiro

Prof. Dr. Ricardo Silva Kubrusly
Universidade Federal do Rio de Janeiro

FOLHA_BRANCA

Para minha mãe, minha primeira
conexão com o mundo e com minha própria
história.

AGRADECIMENTOS

Em um estranho momento de isolamento, afastado de convívio social mais diverso e mergulhado em pensamentos caóticos pela conclusão de mais uma etapa de minha vida profissional e acadêmica, devo agradecer mais uma vez a minha amiga e orientadora Regina Dantas e sua família, que num movimento inverso, agora vêm viver em minha terra, Natal, onde eu pude oferecer apoio e adaptação, assim como ela o fez quando vivi no Rio de Janeiro, seu local de origem.

Agradeço a minha colega de curso Ana Paula Gonçalves também pelo apoio à distância, nossas conversas sobre a relação entre Matemática e História serão sempre um condutor para minhas leituras sobre Redes. Agradecimento especial também ao querido Paulo Aprígio, amigo, colega de curso e de campo, onde a História será sempre um laço que nos liga entre Natal e o Rio de Janeiro. Agradecimento ao amigo André Campos da Rocha que mantem sempre a energia de lutar por uma ciência justa e combativa, sem nenhum passo para trás.

Agradeço aos Professores Eduardo Nazareth Paiva, Ricardo Kubrusly, Rundsthen Nader e Nadja Paraense pela amizade, irreverência e competência nas relações estabelecidas ao longo dos anos de pós-graduação.

Por fim, agradeço novamente à minha mãe, pelo retorno a terra natal, e que em meio a uma batalha contra um câncer e o enfrentamento da pandemia de COVID-19 continua forte em seu apoio e cuidado incondicional para que eu consiga trilhar o meu caminho para autonomia e conquista profissional.

O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001.

“Não vamos fingir que os fenômenos mentais podem ser mapeados nas características das bolas de bilhar.”

Gregory Bateson

RESUMO

MAIA, Daniel. Humanidades Digitais: uma Emergência Teórico-Metodológica na Representação e Descoberta de Conhecimento para a História das Ciências no Brasil. Rio de Janeiro, 2021. Tese (Doutorado em História das Ciências, das Técnicas e Epistemologia) – Programa de Pós-Graduação em História das Ciências, das Técnicas e Epistemologia/HCTE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

A presente pesquisa apresenta uma proposta teórico-metodológica enquanto política científica para análise documental textual como parte do desenvolvimento do campo transdisciplinar das Humanidades Digitais em diálogo com a História das Ciências no contexto brasileiro. Tem como objetivo propor novos modos de representação e descoberta de conhecimento a partir de uma "metaescrita", com a utilização de *softwares* de processamento de texto para a construção de uma fonte híbrida computacional-imagética. Apresentando o estado da arte das humanidades digitais, exploramos o desenvolvimento de estudos transdisciplinares no uso de técnicas de Inteligência Artificial e Aprendizagem de Máquina para Análise Textual do periódico *Archivos do Museu Nacional* e a constituição de redes epistêmicas complexas a partir da análise das narrativas e a formação de banco de dados para elaboração de cartografias processuais. Estabelecendo diálogos entre Manuel Castells e Albert-László Barabási com o conceito de Redes e as novas perspectivas na interação humano-máquina que se apresentam com novos atores das Humanidades Digitais como Melissa Terras, Lisa Gitelman e Stephen Ramsay vemos surgir uma nova epistemologia para as Humanidades na busca pela Representação e Descoberta de Conhecimento. Como resultado, observamos um contexto complexo na implementação de uma política científica documental que permita superar as dificuldades do campo da História das Ciências no que diz respeito ao armazenamento, preservação e disseminação dos saberes e artefatos científicos e históricos. No entanto, são promissoras as perspectivas teórico-metodológicas para o avanço dos diálogos e transformações por meio da Transdisciplinaridade, tanto na formação dos Humanistas como na inovação dos campos de pesquisa, seja com artefatos e documentos físicos ou as novas mídias e textos digitais.

Palavras-chave: História das Ciências; Transdisciplinaridade; Humanidades Digitais.

ABSTRACT

MAIA, Daniel. Digital Humanities: a Theoretic-Methodological Emergence in Knowledge Discovery and Representation for the History of Science in Brazil. Rio de Janeiro, 2021. Thesis (Doctorate in History of Sciences, Techniques and Epistemology) – Post-graduate Program in History of Science, Techniques and Epistemology/HCTE, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2021.

This research presents a theoretical-methodological proposal as a scientific policy in textual document analysis as part of the development of the transdisciplinary field of Digital Humanities in dialogue with the History of Sciences in the Brazilian context. Its objective is to propose new ways of representing and discovering knowledge from a “meta-writing”, with the use of word processing *software* for the construction of a hybrid computational-image source. Presenting the state of the art of digital humanities, we explore the development of transdisciplinary studies in the use of Artificial Intelligence and Machine Learning techniques for Textual Analysis of the journal Archivos do Museu Nacional and the constitution of complex epistemic networks from the analysis of narratives and the formation of a database for the elaboration of procedural cartographies. Establishing dialogues between Manuel Castells and Albert-László Barabási with the concept of Networks and the new perspectives on human-machine interaction presented by new actors in Digital Humanities such as Melissa Terras, Lisa Gitelman and Stephen Ramsay, we see a new epistemology for the Humanities emerging in the search for Knowledge Representation and Discovery. As a result, we observed a complex context in the implementation of a documental scientific policy that allows overcoming the difficulties in the field of History of Sciences with regard to the storage, preservation and dissemination of scientific and historical knowledge and artifacts. However, the theoretical-methodological perspectives for the advancement of dialogues and transformations through Transdisciplinarity are promising, both in the training of Humanists with computing methods and techniques, and in the innovation of research fields, whether with physical artifacts and documents or the new media and digital texts.

Keywords: History of Science, Transdisciplinarity, Digital Humanities.

LISTA DE ILUSTRAÇÕES

FIGURA 1: INTERFACE DO *OBSIDIAN* E SEUS CAMPOS DE INTERAÇÃO.....19
FIGURA 2: REPRESENTAÇÃO IMAGÉTICA EM GRAFOS DE CONCEITOS TEXTUAIS FORMATADOS EM MARKDOWN 20
FIGURA 3: INTERFACE DO *OBSIDIAN* APRESENTANDO SEÇÕES ESTRUTURADAS COM FUNÇÕES DISTINTAS.....21
FIGURA 4: MAPEAMENTO DA INTERAÇÃO ENTRE SOCIEDADE E COMUNIDADE CIENTÍFICA EM PAÍSES EUROPEUS.....24
FIGURA 5: REPRESENTAÇÃO DAS PRODUÇÕES CIENTÍFICAS DO PERIÓDICO ARCHIVOS DO MUSEU NACIONAL.....25
FIGURA 6: APLICAÇÃO DE "THRESHOLD" COMO CONTRASTE ENTRE PRETO E BRANCO, UM PRÉ-PROCESSAMENTO DE IMAGEM ANTES DO OCR.....28
FIGURA 7: RECONHECIMENTO DO CARACTERE ESPECIAL DO LATIN "Æ", A PARTIR DE PADRÃO DE RECONHECIMENTO INTERNO DO ABBYY FINEREADER 15. 29
FIGURA 8: PROCESSAMENTO PARCIAL DE CARACTERES RESULTANDO EM ERROS DE GRAFIA E AUSÊNCIA DE COMPARATIVO COM DICIONÁRIO INTERNO, DEVIDO DIFERENÇAS COM O PORTUGUÊS ARCAICO. 30
FIGURA 9: IDENTIFICAÇÃO DE ENTIDADES NOMEADAS NA CATEGORIA "PESSOAS". 33
FIGURA 10: REPRESENTAÇÃO DE TEMAS E CAMPOS DO CONHECIMENTO CARTOGRAFADOS A PARTIR DE REVISÃO BIBLIOGRÁFICA.....41
FIGURA 11: OCORRÊNCIAS DO TERMO “DIGITAL HUMANITIES” - PUBLICAÇÕES/ANO - INDEXAÇÃO SCOPUS 49
FIGURA 12: PUBLICAÇÕES DO PERIÓDICO “COMPUTERS AND THE HUMANITIES” (1966-2004). 51
FIGURA 13: RECORTE DE DOCUMENTO DESTACANDO ELEMENTOS QUE APRESENTAM DIFICULDADES NO PROCESSAMENTO DO OCR, POR SE TRATAREM DE MANCHAS, GRÁFICOS OU TIPOGRAFIAS INCOMUNS. 68
FIGURA 14: RECORTE DE DOCUMENTO DESTACANDO ELEMENTOS RECONHECIDOS E DIFERENCIADOS DE MODO AUTOMÁTICO PARA PROCESSAMENTO DOS CARACTERES E DEMAIS OBJETOS GRÁFICOS.....69
FIGURA 15: PADRÕES TIPOGRÁFICOS CONSTRUÍDOS A PARTIR DO ABBYY FINEREADER NO PROCESSO DE OCR DO PERIÓDICO ARCHIVOS DO MUSEU NACIONAL V.I (1876). 72

FIGURA 16: GRAFO DE ENTIDADES NOMEADAS, INFORMAÇÃO EXTRAÍDA DO VOL. I DE 1876 DO PERIÓDICO ARCHIVOS DO MUSEU NACIOANAL. 75
FIGURA 17: CARTOGRAFIA DAS RELAÇÕES CIENTÍFICAS NO PERIÓDICO ARCHIVOS DO MUSEU NACIONAL 79
FIGURA 18: INTERFACE DO *OBSIDIAN* COM HASHTAGS E CATEGORIAS DE ENTIDADES NOMEADAS.....80
FIGURA 19: MICRORRELAÇÕES ENTRE QUATRO ENTIDADES NOMEADAS. 81
FIGURA 20: CARTOGRAFIA DAS PERMUTAS DOS ARCHIVOS DO MUSEU NACIONAL.....83

SUMÁRIO

LISTA DE ILUSTRAÇÕES 10

SUMÁRIO 12

1 INTRODUÇÃO 1

1.1. A TEMÁTICA DAS HUMANIDADES DIGITAIS 2

1.2. OS LIMITES DA PESQUISA 3

1.3. CARTOGRAFANDO AS PROBLEMÁTICAS DA PESQUISA 5

1.4. HIPÓTESES PROPOSTAS 6

1.5. OS OBJETIVOS DA PESQUISA PARA A HISTÓRIA DAS CIÊNCIAS 6

1.6. JUSTIFICATIVAS DA PESQUISA 7

1.7. FUNDAMENTAÇÃO TEÓRICA 9

1.8. METODOLOGIAS DA PESQUISA 12

1.9. APRESENTAÇÃO DOS CAPÍTULOS 14

2 DIMENSÃO METODOLÓGICA: POLÍTICAS DE TRANSGRESSÃO NA REPRESENTAÇÃO E DESCOBERTA DE CONHECIMENTO 16

2.1 O *SOFTWARE OBSIDIAN* COMO METAESCRITA 17

2.2 A REPRESENTAÇÃO DO CONHECIMENTO A PARTIR DA TEORIA DE GRAFOS E DA TEORIA DE REDES COM USO DE *SOFTWARES* 22

2.3 PROCESSOS COMPUTACIONAIS NO TRATAMENTO DE DOCUMENTAÇÃO TEXTUAL 26

2.3.1 *Aquisição de Dados em Documentos: o Reconhecimento Óptico de Caracteres (OCR)* 26

2.3.2 *Limpeza de dados: o Proofreading, quando a computação não basta* 29

2.3.3 *Exploração de dados: procurando as pistas para análise documental. Reconhecimento De Entidades Nomeadas (N.E.R.) e Extração de Entidades Nomeadas* 31

2.3.4 *Modelagem de dados: reconhecimento de padrões e replicabilidade analítica* 33

2.4 POR UMA METODOLOGIA CRÍTICA EM HUMANIDADES DIGITAIS 36

3 DIMENSÃO TEÓRICA: A EMERGÊNCIA TEÓRICO-METODOLÓGICA PARA A HISTÓRIA DAS CIÊNCIAS 38

3.1 OS EPICENTROS TEÓRICOS TRANSDISCIPLINARES PARA AS HUMANIDADES DIGITAIS 41

3.2 O *ETHOS* DAS HUMANIDADES DIGITAIS: PENSAR, FAZER, SER 47

3.3 O DIGITAL: UM (NEM TÃO NOVO) TERRITÓRIO DAS HUMANIDADES 49

3.4 O DOCUMENTO COMO TERRITÓRIO: VAMOS CARTOGRAFAR! 61

Cartografia como Pesquisa-Intervenção 61

A Atenção do Cartógrafo 62

Cartografar é Acompanhar Processos 62

O Coletivo de Forças como Plano da Experiência Cartográfica 63

Cartografar é Habitar um Território Existencial 63

Por uma Política da Narratividade 64

4 DIMENSÃO EXPERIMENTAL: REPOSITÓRIOS CIENTÍFICOS E SUAS POTENCIALIDADES METODOLÓGICAS 66

4.1 O TERRITÓRIO DO HISTORIADOR FRENTE OS DOMÍNIOS DAS HUMANIDADES DIGITAIS 66

4.2 O PRIMEIRO CONTATO: NAVEGANDO PELOS *ARCHIVOS DO MUSEU NACIONAL* 68

4.3 OCR E DOCUMENTAÇÃO HISTÓRICA: O RUÍDO FAZ A DIFERENÇA 70

4.3.1 *Reconhecimento de Padrões Tipográficos e a Construção de um Banco de Dados Linguístico* 72

4.3.2 *Transgressões da forma: close reading, distant reading e algo além.* 75

4.4 CARTOGRAFIAS DO MUSEU NACIONAL: O ESPAÇO DE FLUXOS DAS CIÊNCIAS NO BRASIL DO SÉC. XIX SOB A ÓTICA DAS HUMANIDADES DIGITAIS NO SÉC. XXI 79

CONSIDERAÇÕES FINAIS (A FUGA DO GATO DE SCHRÖDINGER) 88

REFERÊNCIAS 94

ANEXO 1 102

ANEXO 2 103

ANEXO 3 107

ANEXO 4 110

1 INTRODUÇÃO

Dando continuidade à experiência acadêmica, no Programa de Pós-graduação em História das Ciências da Universidade Federal do Rio de Janeiro, me vi envolvido pelos estudos transdisciplinares que contemplavam abordagens históricas, epistemológicas e das tecnologias humanas como uma premissa para **a proposição de novas políticas científicas no tratamento de documentos históricos**.

A pesquisa de mestrado, defendida anteriormente (MAIA, 2017), surgiu pela compreensão da interação sistemática entre múltiplos métodos e teorias científicas em busca de uma análise histórica mais ampla sobre as ciências no Brasil a partir de experimentações técnicas em documentos históricos, resultando não só numa fonte escrita, mas também num novo tipo, complexo e multifacetado, um grafo computacional-imagético.

A partir de estudo de caso, realizado com a documentação do periódico "*Archivos do Museu Nacional*", foi possível identificar um agenciamento das redes institucionais científicas no período que correspondeu ao fim do Império do Brasil e a Proclamação da República. Ao propor uma metodologia de visualização de dados históricos é possível uma leitura não-linear e imagética das informações textuais, abrindo um caminho para novos modos de racionalidade sobre a narrativa histórica documental.

Portanto, foi possível constatar a viabilidade da aplicação teórico-metodológica com um viés transdisciplinar ao usar a Teoria de Redes, retomado na presente pesquisa de doutorado, em conjunto com os conceitos de Aprendizagem de Máquina, Inteligência Artificial e Representação do Conhecimento, sob o termo Humanidades Digitais, respondendo às questões iniciais sobre a constituição de redes institucionais científicas e de sujeitos. Deste modo, reconheço a aplicabilidade desse modelo transdisciplinar de análise para além da ciência histórica.

Nessa perspectiva, tornou-se necessário apontar as idiossincrasias inerentes à pesquisa, dificuldades técnicas e insuficiências teóricas, que por mais que tenham sido inspiração para críticas no campo da Epistemologia e da Metodologia das

Ciências, fazem parte de um longo processo de aprofundamento da pesquisa, cujo tempo se estende para além do contexto formativo da pós-graduação.

Com o objetivo de constituir os pilares teórico-metodológicos da pesquisa, um conjunto de autores e conceitos presentes desde a Sociologia das Ciências e a Ciência da Informação até a Computação e a Representação do Conhecimento com uso de *softwares* especializados, são apresentados no texto com o que chamei de Dimensão Teórica, servindo de guia para a formulação dos argumentos que permeiam os demais capítulos em defesa da metáfora das redes nas análises metodológica e política da emergência das Humanidades Digitais.

Com essa abstração das redes em comum é possível elaborar algumas relações entre as abordagens teóricas, questionando os limites entre os campos de conhecimento, entre os métodos adotados, extraindo e analisando o conteúdo a partir de documentação histórica, sem ignorar as pesquisas da atual historiografia da História das Ciências, e numa abordagem prática experimental, a utilização de *software* de computador na transcrição das informações coletadas.

Um triplo esforço de reconhecimento das redes de agenciamento a partir de documentos que revelam outras perspectivas para as narrativas sobre as relações científicas institucionais como representação e descoberta de conhecimento.

1.1. A TEMÁTICA DAS HUMANIDADES DIGITAIS

Não é de hoje que as civilizações tentam entender a si mesmas se utilizando de saberes e técnicas. O século XIX tem merecido o crédito de ter sido berço temporal das Humanidades enquanto disciplinas e a linha do tempo que se segue desde então tem dado lugar às tecnologias como modo de operacionalização de uma racionalidade mais globalizada.

Cabe aqui tentar entender então uma amalgama emergente entre as Humanidades e a Tecnologia, que vem descortinando potencialidades analíticas e ampliando nossas capacidades gerais de construção, partilha e preservação de conhecimento.

As Humanidades Digitais são um campo emergente da disciplinarização do conhecimento, mais especificamente, uma transdisciplinarização. Haja vista a essência de seu arcabouço teórico-metodológico ser multifacetado. Servindo como um “termo guarda-chuva”, as Humanidades Digitais têm atuado em grandes áreas das Artes, da Medicina, da Computação e das próprias Ciências Humanas e Sociais, onde o momento de seu desenvolvimento gira em torno do contexto da Segunda Guerra Mundial, com o avanço das comunicações globais e a miniaturização tecnológica.

O senso comum nos leva a pensar que as Humanidades Digitais são métodos computacionais que tentam entender o que é ser humano em sociedades do passado e do presente. E uma das coisas mais “fáceis” de fazer por meio dos computadores é contabilizar coisas, mas para isso é necessário transformar tais coisas numa linguagem que possa ser interpretada pelo computador, ou seja, binarizar. Assim, a digitalização, processo de transcrição de um objeto físico para o meio digital, é tida como o principal processo de mudança nas metodologias quantitativas para as humanidades. (TERRAS, 2013).

1.2. OS LIMITES DA PESQUISA

Com sua característica fundacional recente e em constante construção, o tema das Humanidades Digitais vem tentando adaptar métodos tradicionais de investigação científica, filosófica e social para sanar as dificuldades contemporâneas das ciências que, por ainda adotarem paradigmas teóricos hierarquizantes, não têm conseguido resolver os problemas complexos que ora se apresentam, a exemplo da suposta separabilidade entre humanidades, as tecnologias e a natureza. Para os pesquisadores dessa crescente área é, e talvez continue sendo por muito tempo, inevitável o confronto com tal paradigma.

Sendo uma pesquisa no campo da História das Ciências e das Tecnologias, há a oportunidade de avaliar fontes históricas textuais a partir da utilização de ferramentas computacionais, o que nos leva a construção de fontes híbridas, binarizadas, de tal modo que nos permita uma proposição teórico-metodológica

transdisciplinar como política científica de análise documental para a História das Ciências no Brasil.

O caráter transdisciplinar da pesquisa abarca discussões na esfera teórica sobre as fronteiras epistemológicas das Humanidades Digitais, também o seu potencial metodológico ao congregar múltiplos modos de análise e por fim um caráter experimental, reflexo da emergência do tema na busca por soluções para os complexos problemas de ordem social e política na construção de conhecimento ainda no século XXI.

Nos encontramos com as demandas da disciplina histórica agora no contexto do século XXI, e novos problemas se apresentam à medida que novas tecnologias surgem. Instituições governamentais e científicas, que lidam com preservação e manutenção de suas fontes históricas escritas, têm implementado ao longo dos últimos 30 anos, projetos de digitalização de seus acervos[1]. Isso abriu caminho para novas perspectivas teórico-metodológicas no campo das Humanidades Digitais.

Nessa perspectiva, para a presente investigação, se fez uso de *software* específico na construção de um novo método de escrita e visualização de informações para a representação do conhecimento, ainda que saibamos das indeterminações de uma abordagem transdisciplinar, e mesmo de capacidade técnica, em que pese a inovação tecnológica com modelo de código aberto e de participação de sujeitos não-especialistas.

Dessa forma, foi possível alimentar os *softwares* com dados extraídos de documentos históricos do século XIX escaneados, onde os originais, preservados na Seção de Memória e Arquivo do Museu Nacional/SEMEAR, não foram acessados a tempo, devido o trágico incêndio do Museu Nacional da UFRJ no dia 2 de setembro de 2018. Tal fato levou a uma limitação na consulta de fontes complementares em potencial, como a documentação administrativa da instituição[2].

---

[1] Tais como os projetos da Biblioteca Nacional e o da Biblioteca Brasiliana. Ver mais em: <http://bndigital.bn.gov.br/> e https://www.bbm.usp.br/pt-br/.

[2] O evento catastrófico do incêndio no Museu Nacional ao mesmo tempo que destruiu todo o acervo que se localizava no prédio principal fez revelar também, escondida entre os destroços, a antiga estrutura em alvenaria, como também as pinturas originais. Um evento como esse não permite fechar a tempo as feridas e lacunas institucionais no que diz respeito às políticas de preservação, mas foi um momento de luta coletiva e engajamento para a elaboração de novos protocolos de recuperação da

Outras limitações de ordem técnica dizem respeito ao conhecimento de linguagem de programação, não totalmente dominada, haja vista a formação tradicional como historiador, o que abre caminho para discussões críticas acerca da reformulação dos currículos em humanidades no século XXI.

1.3. CARTOGRAFANDO AS PROBLEMÁTICAS DA PESQUISA

Os processos de identificação das questões referentes à pesquisa dão continuidade à pesquisa anterior, de mesmo tema e abordagem. Enquanto pesquisador da área das ciências humanas, mais especificamente a História, como podemos nos utilizar de métodos quali-quantitativos para análise documental mediadas por computador? A lacuna na formação transdisciplinar dos humanistas é uma questão séria a ser debatida também aqui enquanto proposta metodológica de política científica.

Estamos chamando a atenção para uma linguagem incomum, que pode atravessar os diversos e, muitas vezes opostos, campos do conhecimento científico, como também para métodos de análise que identifiquem a complexidade das esferas social e política na construção do conhecimento científico e a percepção que a sociedade tem deste.

Assim, a questão que move a pesquisa é compreender como a adoção de múltiplos métodos de análise documental pode revelar novos caminhos para a representação e descoberta de conhecimento histórico e científico. É possível para historiadores das Ciências, pela mediação da tecnologia, um novo fazer historiográfico e científico?

---

estrutura física do Museu e de inovação e colaboração, haja vista a participação de diversas instituições museológicas ao redor do mundo oferecendo parte de seus acervos para a reconstituição do que foi perdido no Museu Nacional. No que diz respeito aos protocolos de preservação, publicação e acesso ao acervo digital institucional, esta tese traz à luz algumas estratégias de reconfiguração do fazer científico e historiográfico para a documentação, em busca de facilitar e acelerar as pesquisas em suas bases de dados pré-existentes. Equipamentos e servidores locais, que se encontravam em diversas salas do prédio do Museu, foram destruídos, assim como seus arquivos. É essa reconfiguração necessária a qual chamamos atenção no âmbito das políticas científicas tendo as Humanidades Digitais como guia.

1.4. HIPÓTESES PROPOSTAS

A emergência do campo das Humanidades Digitais tem mudado consideravelmente a cultura científica de preservação e análise documental, fornecendo novas ferramentas, métodos e teorias. Tem nos permitido a visualização e descoberta de antigas, e mesmo novas, relações históricas entre sujeitos, instituições e saberes científicos.

A partir da experimentação metodológica proposta, espera-se que a utilização de *software* especializado para o processamento de documentação escaneada proporcione maior celeridade no tempo de pesquisa de base textual, como também novos modos de representar e visualizar conhecimento, abrindo caminho para novas narrativas científicas.

1.5. OS OBJETIVOS DA PESQUISA PARA A HISTÓRIA DAS CIÊNCIAS

O objetivo geral da presente pesquisa se dá a partir da proposição de metodologia transdisciplinar, inserida no contexto emergente do campo das Humanidades Digitais, como modo de saber e fazer ciências na busca por novas formas de Representação e Descoberta de Conhecimento.

Para tanto, procuramos: propor uma abordagem de extração de dados de documentação textual para a representação de conhecimento histórico e científico por meio de processos computacionais-imagéticos com *softwares* especializados; apresentar, como dimensão teórica, o estado da arte das Humanidades Digitais e sua relação com as práticas metodológicas na História enquanto disciplina; experimentar a proposição metodológica num estudo de caso a partir do periódico "*Archivos do Museu Nacional"*.

1.6. JUSTIFICATIVAS DA PESQUISA

A motivação para o desenvolvimento dessa proposta de pesquisa é resultado das transformações políticas e científicas no campo das Humanidades no contexto brasileiro. Em consequência do incêndio do Museu Nacional da UFRJ, em setembro de 2018, a possibilidade de pesquisa se restringiu à documentação digitalizada remanescente. A partir disso, reflexões teórico-metodológicas foram incorporadas ao objetivo inicial de analisar as potencialidades da construção de novas fontes primárias a partir de documentação histórica com utilização de ferramentas computacionais.

Ademais, verificou-se uma lacuna no contexto nacional relacionada ao tema das Humanidades Digitais. Temos, assim, a oportunidade de desenvolvê-lo com a especificidade das demandas locais e propor diretrizes para as políticas científicas que possam servir de modelo para instituições que apresentem demandas semelhantes.

Um dos primeiros obstáculos observados na análise documental, levando-se em consideração a preocupação com a eficácia da extração dos dados, é a impossibilidade da automação completa da identificação dos termos sensíveis na elaboração de um banco de dados relacionais.

No trabalho historiográfico, a análise documental representa uma grande parcela do tempo dedicado à pesquisa. Ainda assim, tendo hoje disponível tecnologia capaz de reconhecer padrões de escrita, conhecida como OCR (*Optical Character Recognition*), tal recurso não pôde ser utilizado com eficácia devido o tempo que levaria para corrigir a imprecisão do próprio procedimento, fazendo com que fosse reavaliado para a presente pesquisa.

Isso gerou uma dificuldade na precisão da coleta de dados e a própria construção das imagens, visto que o *software* poderia ser alimentado de forma automática com o output do processo de reconhecimento de caracteres. Em conformidade com as condições de coleta e extração de dados, a pesquisa prossegue também com uma bibliografia complementar de origem eletrônica, a partir de livros e artigos encontrados na Internet.

Isso me trouxe mais questionamentos para a pesquisa no que diz respeito não só a elaboração de novos métodos e sobre a criatividade individual do cientista, mas também sobre as políticas científicas elaboradas no seio das instituições governamentais e das comunidades acadêmicas, caracterizado pelo que Price e Kuhn afirmaram como “pequeno-mundo” e os custos financeiros restritos a um modelo de política que não proporciona acesso amplo à informação, à própria comunidade científica e ao público em geral (PRICE, 1963; KUHN, 1962).

Ainda com uma perspectiva dicotômica e dualista, a proposta por uma ciência de acesso aberto tem seus detratores e defensores. Os primeiros argumentam sobre o potencial mal-uso dos dados, quando coletados por um indivíduo e aproveitado sem esforço por outro. Argumentam também que uma enorme quantidade de dados dispersos sobrecarregaria a comunidade científica, uma questão que pode ser resolvida com métodos de filtragem oferecidos pela própria área da ciência dos dados com o uso das métricas. Por outro lado, a favor de uma ciência aberta, estaria um número cada vez mais crescente de cientistas valorizando a transparência das políticas científicas na validação das descobertas quando toda a comunidade tem acesso aos dados.

Assim como uma antiga questão fundacional sobre a propriedade do conhecimento humano ser legado à esfera pública e não ao domínio privado, em se tratando de financiamento público, o conhecimento científico deve ser considerado também de domínio público e livre em seu acesso. É com base nessa premissa que a presente pesquisa continua com uma diretriz ética na elaboração de uma metodologia transdisciplinar, com potencial de análise aberta a vários campos do conhecimento e seus sujeitos, não apenas a História.

É com tamanha complexidade que as Humanidades Digitais vêm emergindo, tendo em seu conjunto ferramental uma forte influência da teoria de redes, a qual atravessa diversos campos de saberes. A natureza transdisciplinar das Humanidades Digitais oferece uma linguagem pela qual diferentes disciplinas podem interagir de modo inovador umas com as outras. É possível que historiadores e cientistas da computação se vejam confrontados com a mesma tarefa de elaborar os diagramas de agenciamento por trás de seus sistemas de pensamento, extraindo informações de

conjuntos de dados às vezes incompletos e ruidosos, e a necessidade de entender sua robustez pela emergência de novas metodologias transdisciplinares.

1.7. FUNDAMENTAÇÃO TEÓRICA

Seguindo a proposição da presente pesquisa, faz-se necessário uma breve abordagem sobre os autores que dão base às teorias e métodos aqui propostos. Importante apontar uma premissa na escolha de cada um deles: o conceito de Rede, e suas metáforas de agenciamento, como parte das epistemologias de fronteira, construindo uma cartografia teórico-metodológica na elaboração do que chamo aqui de Rede Epistêmica Complexa.

Manuel Castells

Tomando de empréstimo os estudos de Manuel Castells, adotamos aqui a perspectiva de alguns conceitos elaborados em "A Sociedade em Rede" como dimensão que constitui o espaço de fluxos das relações científicas institucionais ao analisar a documentação do Museu Nacional identificando sujeitos, instituições e ideias que circulavam por meio do periódico Archivos do Museu Nacional como um registro histórico da atividade científica no contexto brasileiro. Com a transformação dos meios de comunicação em massa tradicionais em direção a um modo de operação horizontal e descentralizado no contexto da Internet, uma nova linguagem, binarizada e manipulável, foi introduzida na interpretação da realidade social e nos padrões de comunicação, não só para colocar em xeque os meios tradicionais de verificação da verdade, como também descortinar narrativas ocultas em entrelinhas que só agora se tornaram possíveis com a interação humano-máquina.

A relevância de uma nova linguagem do mundo digital, para explicar a evolução social civilizacional e a análise dos registros históricos e do conhecimento, é parte integrante da presente proposta teórico-metodológica, de caráter transdisciplinar. O conceito de Espaço de Fluxos desenvolvido por Castells é exemplo que constitui essa nova linguagem caracterizada pela metáfora da rede em que é utilizada enquanto categoria analítica, na compreensão das representações em clusters na transdução

de fontes primárias analógicas em fontes digitais como um processo de construção de novas narrativas visuais do conhecimento histórico e científico.

Derek John de Solla Price

Considerado o criador da Cientometria, seus trabalhos envolvem estudos estatísticos das dinâmicas produtivas dos campos do conhecimento científico. Tendo elaborado estudos quantitativos das redes de citações entre escritos acadêmicos, Price contribuiu para a identificação do que se entende por rede de escala livre, sendo assim, a adoção de Price como um dos autores base, principalmente com a obra Little Science, Big Science...and Beyond (1986), nos serve para a compreensão do cenário macro e micro institucional da produção de conhecimento histórico e científico a partir de análise do periódico Archivos do Museu Nacional adotando metodologia transdisciplinar para uma modelagem do referido cenário.

Thomas S. Kuhn

Ciência Normal, Crises, Anomalias, Fronteiras e Paradigmas são alguns dos conceitos trabalhados por Kuhn na identificação das atividades e produções das comunidades científicas. Em consonância com a cientometria, a utilização de tais conceitos para a compreensão das redes epistêmicas constituídas a partir de análise textual representa um processo indispensável da presente pesquisa na construção de uma cartografia das relações científicas institucionais e epistemológicas. É a partir desse processo que poderemos compreender a formação e desenvolvimento dos diversos campos das ciências no Brasil no séc. XIX, a exemplo da Antropologia, Geologia e Botânica, três dos principais campos desenvolvidos no Museu Nacional e sendo tema central do periódico Archivos do Museu Nacional desde sua primeira publicação.

Niklas Luhmann

Suas teorias sistêmicas englobam tanto os campos da biologia e da psique, mas também, e principalmente o campo da sociologia. A formulação de suas ideias se aproxima de modo mais geral do conceito de complexidade, ao invés das grandes

teorias unificadas, objeto de crítica da pós-modernidade. A ideia de sistemas sociais desenvolvidos por Luhmann visa contextualizar toda vida social dentro de um modelo teórico universal, uma modelagem complexa como um “labirinto”, ou seja, um modelo de análise não-linear. Luhmann considera os sistemas sociais eminentemente como sistemas de comunicação, com fronteiras definidas por si mesmos e pelo ambiente. Comunicações internas a um sistema ocorrem a partir de uma limitada seleção de informações externas filtradas por um conjunto de significados, sendo tal processo denominado de redução de complexidade. A compreensão de um conjunto de elementos em agenciamento sistemático constitui a metáfora analítica da rede epistêmica proposta na presente pesquisa por meio da interface de *software* especializado. Sendo cada comunidade científica, representada por indivíduos, instituições ou publicações, os sistemas micro e macro numa cartografia processual da atividade científica no Brasil extraída a partir da análise documental do periódico Archivos do Museu Nacional transcrito para o meio digital.

Mark Granovetter

O trabalho de Granovetter serve de base para a presente pesquisa no que tange as complexas relações de agenciamentos entre entidades das redes institucionais científicas e o valor de suas influências a partir das conexões estabelecidas com outras entidades. Atravessando os campos da Ciência da Informação, Sociologia das Ciências e a Antropologia, o conceito de “Força dos Laços Fracos” permite análises de sujeitos ou grupos isolados dentro de uma rede complexa revelando sistemas menores dentro de sistemas maiores, e no que diz respeito a informação textual, pode revelar a construção de relações semânticas antes invisíveis diante de uma disposição linear tradicional.

Basarab Nicolescu

A partir do “campo” da Transdisciplinaridade (ANEXO 4), os escritos de Nicolescu permeiam a presente pesquisa com o objetivo de revelar os diversos, e muitas vezes opostos, níveis de realidade que coexistem nas redes epistêmicas reveladas a partir do estudo dos espaços de ciência e de suas instituições, sujeitos e

documentos. Não estando limitada a reconhecer a reciprocidade dos campos do conhecimento, a transdisciplinaridade procura localizar e cartografar processos inerentes aos micro e macro sistemas sem definir fronteiras ontológicas e epistemológicas definitivas, levando o exercício do pensar epistemológico para além da disciplinarização dos saberes.

1.8. METODOLOGIAS DA PESQUISA

A pesquisa se caracteriza pelo estudo bibliográfico com a utilização de método quali-quantitativo e segue na construção de uma metodologia transdisciplinar que congrega abordagens da Ciência da Informação, da Computação e da Sociologia das Ciências como pistas para a elaboração da rede de agenciamentos entre os sujeitos, instituições e conceitos contidos no conjunto de fontes textuais analisadas em busca de novos modos de representar o conhecimento histórico e científico.

A partir do processo de binarização dos textos escaneados do periódico Archivos do Museu Nacional, enquanto ESTUDO DE CASO (YIN, 2018), a extração das informações e sua organização como um conjunto de grafos, utilizo *softwares* de computador como parte de experimentações na construção de novas fontes híbridas de informação, como grafos e arquivos *markdown*, contendo dados que identifiquem particularidades temporais, geográficas, semânticas ou epistêmicas referentes à atividade científica descritas no periódico do Museu Nacional no contexto da passagem do século XIX para o século XX.

A exemplo do *ABBYY FineReader*, *software* especializado em digitalização e Reconhecimento Óptico de Caracteres (OCR), que permite a binarização de padrões tipográficos, que para a atual pesquisa limita-se às letras de máquina, como prensas mecânicas e mídias do século XIX; também há o *software Gephi*, uma ferramenta para construção de grafos a partir de arquivos ".csv" contendo palavras-chave e suas possíveis correlações; outro *software* é o *Voyant*, uma plataforma online de análise de texto que permite a visualização de dados correlacionados a partir de um corpus documental e banco de dados textuais.

E por fim, como proposta de uma “metaescrita”, apresento aqui a transcrição e visualização de trechos de obras históricas, informações bibliográficas, imagens e notas digitais constituídas a partir do *software Obsidian*, que congrega recursos de referência cruzada, permitindo também a visualização em grafos a partir da leitura de notas escritas no formato *markdown*. Ademais, recorri a uma extensa bibliografia digital, adquiridas a partir da *Library Genesis* e, como fonte primária para o estudo de caso, recorri ao próprio site do Museu Nacional da UFRJ que hospeda os arquivos digitais do Periódico “Archivos do Museu Nacional”[3] para contextualizar a formulação desta proposta de pesquisa transdisciplinar.

Dando continuidade à experiência acadêmica, no programa de pós-graduação em História das Ciências da Universidade Federal do Rio de Janeiro, me vi envolvido pelos estudos interdisciplinares que contemplavam abordagens históricas, epistemológicas e das tecnologias humanas como uma premissa para a proposição de novas políticas científicas no tratamento com os documentos históricos.

A pesquisa de mestrado, defendida anteriormente, surgiu pela compreensão sistemática da interação entre múltiplos métodos científicos em busca de uma análise histórica mais ampla sobre a história das ciências no Brasil a partir de experimentações técnicas em documentos históricos, resultando não só numa fonte escrita, mas também num novo tipo, complexo e multifacetado, um grafo computacional-imagético.

Foi possível constatar a viabilidade da aplicação teórico-metodológica com um viés transdisciplinar ao usar a Teoria de Redes, a ser retomada nesta pesquisa, em conjunto com os conceitos de Aprendizagem de Máquina, Inteligência Artificial e Representação do Conhecimento, sob o termo Humanidades Digitais, respondendo às questões iniciais sobre a constituição de redes institucionais científicas e de sujeitos. Deste modo, reconheço a aplicabilidade desse modelo transdisciplinar de análise para além da ciência histórica.

A partir da análise da documentação do periódico Archivos do Museu Nacional, enquanto estudo de caso, foi possível identificar um agenciamento das redes

[3] Primeiro periódico institucional, criado em 1876 pelo ex-diretor Ladislau Netto, contendo pesquisas das áreas do conhecimento desenvolvidas no Museu Nacional.

institucionais científicas no período que correspondeu ao fim do Império do Brasil e a Proclamação da República. Ao propor uma metodologia de visualização de dados históricos é possível uma leitura não-linear e imagética das informações textuais, abrindo um caminho para novos modos de racionalidade sobre a narrativa histórica documental.

Nessa perspectiva, tornou-se necessário apontar as idiossincrasias inerentes à pesquisa, dificuldades técnicas e insuficiências teóricas, que por mais que tenham sido inspiração para críticas no campo da Epistemologia e da Metodologia das Ciências, fazem parte de um longo processo de aprofundamento da pesquisa, cujo tempo se estende para além do contexto formativo da pós-graduação.

Com o objetivo de constituir os pilares teórico-metodológicos da pesquisa, um conjunto de autores e conceitos presentes desde a Sociologia das Ciências e a Ciência da Informação até a Computação e a Representação do Conhecimento com uso de *softwares* especializados, são apresentados no texto com o que chamei de Dimensão Teórica, servindo de guia para a formulação dos argumentos que permeiam os demais capítulos em defesa da metáfora das redes nas análises metodológica e política da emergência das Humanidades Digitais.

Com essa abstração das redes em comum é possível elaborar algumas relações entre as abordagens teóricas, questionando os limites entre os campos de conhecimento, entre os métodos adotados, extraindo e analisando o conteúdo a partir de documentação histórica, sem ignorar as pesquisas da atual historiografia da História das Ciências, e numa abordagem prática experimental, a utilização de *software* de computador na transcrição das informações coletadas. Uma complexa cartografia para o reconhecimento das redes de agenciamento a partir de documentos que revelam outras perspectivas para as narrativas sobre as relações científicas institucionais.

1.9. APRESENTAÇÃO DOS CAPÍTULOS

Inicialmente, exploramos a **dimensão metodológica** a partir da Interação humano-máquina com uma cartografia de processos e interfaces na lida com

documentação histórica. A partir de uma proposição teórico-metodológica que procura indicar novos caminhos para a descoberta e representação de conhecimento, o uso de interfaces de *software* como proposta transdisciplinar de política científica documental vem apresentar novos modos de pensar e fazer ciências com a constituição de novas fontes primárias, revelando as entrelinhas e as múltiplas narrativas que se apresentam a partir de uma leitura não-linear, imagética e computacional.

Nos diálogos de uma **dimensão teórica**, exploramos a constituição de uma cartografia de conceitos e seus interlocutores por meio da metáfora das redes enquanto linguagem comum que atravessa campos do conhecimento de modo transdisciplinar. Estabelecendo os pilares teóricos do campo das Humanidades Digitais, procuramos compreender a constituição das múltiplas metodologias assimiladas e transformadas em resposta a complexidade da dimensão informacional na qual estamos, indivíduos e instituições, mergulhados. Tendo o documento textual como base, a dimensão digital que se apresenta requer um novo olhar e um novo tratamento na interpretação dos registros históricos para elaborarmos medidas garantidoras da memória coletiva encarando o desafio da efemeridade a qual essa nova dimensão está sujeita.

Como parte de uma **dimensão experimental**, procuramos explorar a aplicação de uma proposta transdisciplinar para dar luz às potencialidades de políticas científicas com base nas discussões em Humanidades Digitais para a História das Ciências. A partir de um estudo de caso de documentação histórica trazemos à luz desafios que métodos da computação apresentam na interação com objetos de pesquisa tradicionais das humanidades, revelando antigos problemas do território do historiador tradicional, como também possibilidades de transgressão dos sujeitos, artefatos e conceitos que habitam esse lugar.

2 DIMENSÃO METODOLÓGICA: POLÍTICAS DE TRANSGRESSÃO NA REPRESENTAÇÃO E DESCOBERTA DE CONHECIMENTO

> "É a busca infinita por um método que previne as humanidades de permanecerem cristalizadas tempo suficiente para qualquer um dizer que seus objetos ou métodos finalmente deram tudo de si e falaram tudo que tinham para falar."
>
> James E. Dobson

Ao analisar o desenvolvimento dos saberes e campos científicos, especialmente com o advento do que entendemos por Ciência Moderna, não podemos negar a delimitação das fronteiras ontológicas e epistemológicas, como também a disciplinarização do conhecimento. E como toda delimitação, a complexidade do desenvolvimento científico e dos problemas que dela derivam fazem com que o processo histórico demande por inovação e criatividade.

Assim, o que entendemos como *descoberta e representação de conhecimento* são parte de infinitos conjuntos de engrenagens, fluxos e interações entre humanos, máquinas, métodos, técnicas e teorias que nos levam à possíveis quebras daquela mesma delimitação inaugurada pelas ciências modernas, não como um reparo, um resgate ou reunificação, mas como uma transgressão.

Evidente que vivenciamos rápidas mudanças no desenvolvimento tecnológico, principalmente no campo da computação, e sua aplicação nas humanidades não tem ficado para trás na corrida por inovação, na defesa de novas metodologias com o objetivo de estender as dimensões interpretativas de um vasto conjunto documental ainda inexplorado por meios computacionais.

Conjuntos documentais que, hoje podendo ser chamados *datasets*, oferecem possibilidades de inovação como também de críticas, haja vista o surgimento de inúmeros novos problemas e questões ao considerar um novo conjunto de dados, tipos híbridos, como fontes históricas e científicas; ao contrário de fontes históricas tradicionais, intocadas ou inalteradas em condições ideais de preservação e validação, esses novos conjuntos de dados devem passar por processos de normalização, manipulações e categorizações para que sejam interpretados e analisados por computadores. (DOBSON, 2019. p. x).

Teóricos das Humanidades Digitais, como David M. Barry e Stephen Ramsay, têm proposto a noção de “computacionalidade” como o objeto formal para as humanidades digitais. Assim, adentrando a conceituação ontológica em que traz à tona uma nova época para as humanidades enquanto uma nova dimensão de inteligibilidade histórica (DOBSON, 2019. p. 3). O contexto contemporâneo das tecnologias de informação e comunicação tem evidenciado a complexidade cada vez maior de um pensar e fazer historiográfico que não seja pela mediação de interfaces computacionais entre humanos e máquinas.

2.1 O *SOFTWARE OBSIDIAN* COMO METAESCRITA

A estrutura produz o conteúdo, e o conteúdo produz a estrutura. É desse modo que compreendemos o impacto transformativo da utilização de interfaces computacionais na produção, representação e descoberta de conhecimento.

Na busca pela compreensão do que pode ser uma “metaescrita” nos deparamos com as ideias de Vilém Flusser ao refletir sobre a interação entre humanos e máquinas por meio dos registros escritos. De tal modo que:

> Pode-se deixar a escrita, essa organização dos sinais, por conta das máquinas. O que se tem em mente aqui não são máquinas de escrever, no sentido tradicional da palavra, uma vez que, nesse caso, é ainda um ser humano que organiza em linhas, com um toque nas teclas, de acordo com as regras da escrita, os sinais gráficos distribuídos no teclado. Aqui temos em mente verdadeiras máquinas de escrever (inteligência artificial), que providenciam elas próprias essa organização (FLUSSER, 2011, p. 25).

Por uma coincidência do momento da pesquisa, depois de idas e vindas de leituras que pudessem dar base a uma pista teórica sobre as humanidades digitais e a função da escrita na representação de conhecimento, explorar as teorias de Flusser agora que nos propomos a falar de metodologia, transgride a ordem de sumário apresentado, o que reflete particularidades transdisciplinares, atravessamentos entre teóricos e metodologias enquanto processos cartográficos.

A partir de uma convergência de mídias, fronteiras entre produção, distribuição e consumo de conhecimento (seja este científico, cultural ou histórico) foram rompidas

com o desenvolvimento cada vez mais acelerado das tecnologias de informação e comunicação, democratizando o acesso à informação. E quando Flusser afirma que as percepções históricas do mundo se dariam apenas pela escrita (FLUSSER, 2011), ousamos dizer que somos arremessados para uma dimensão sociotécnica das comunicações digitais, em que algoritmos tem apresentado desafios aos espaços de fluxos informacionais, colocando em xeque as percepções individuais e coletivas sobre o mundo e os saberes sobre este.

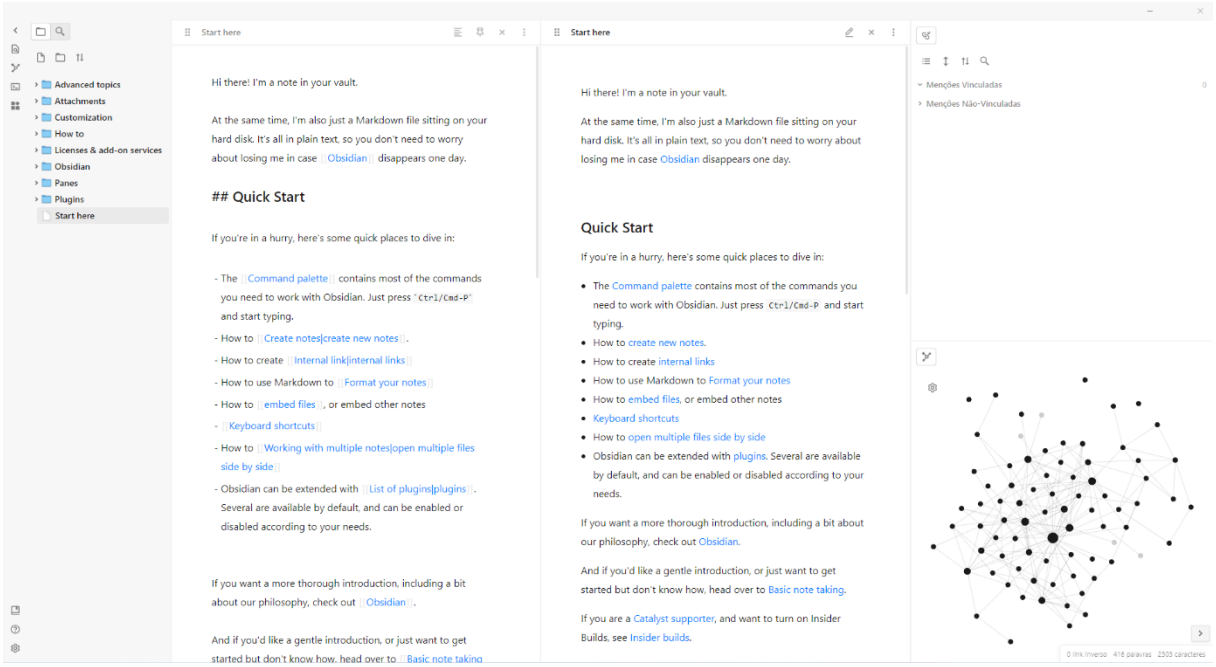
Por outro lado, usando das potencialidades de algoritmos e interfaces digitais, tem sido possível criar novas narrativas na representação e descoberta de conhecimento, descortinando outras percepções de mundo, onde a leitura linear e textual, mesmo na dimensão digital, dá cada vez mais lugar à uma linguagem híbrida, interativa e imagética, onde o rearranjo e manipulação das informações e bancos de dados produzem novos sentidos sobre o passado, o presente e mesmo o futuro dos sujeitos, interconectados e codependentes em definitivo.

E é por meio de um exercício de metaescrita, como proporia Flusser, a partir de uma interface interativa, que exploramos essa dimensão metodológica da pesquisa com o uso do *software* ***Obsidian*** na experiência de uma cartografia processual para o tratamento e análise de documentação histórica.

É possível observar abaixo (Figura 1) a disposição dos elementos da interface do *Obsidian*, de modo que cada seção cumpre uma função específica para a construção e representação de conhecimento. O *software* é dotado de funcionalidades que permitem uma representação imagética de uma escrita tanto linear, como não-linear. Ou seja, ao lado direito da interface vemos uma representação em grafos, uma rede de pontos interconectados que traduzem em uma imagem a relação entre entidades textuais apresentadas de um modo linear na parte central da interface.

Diferentemente de processadores de texto tradicionais, como o *Microsoft Word*, o *Obsidian* tem o propósito de representar a escrita de maneira atomizada, por meio de arquivos individuais dotados de códigos textuais variados, sejam anotações, cópias de textos originais, expressões computacionais, referências a outros arquivos, etc.

Figura 1: Interface do *Obsidian* e seus campos de interação.



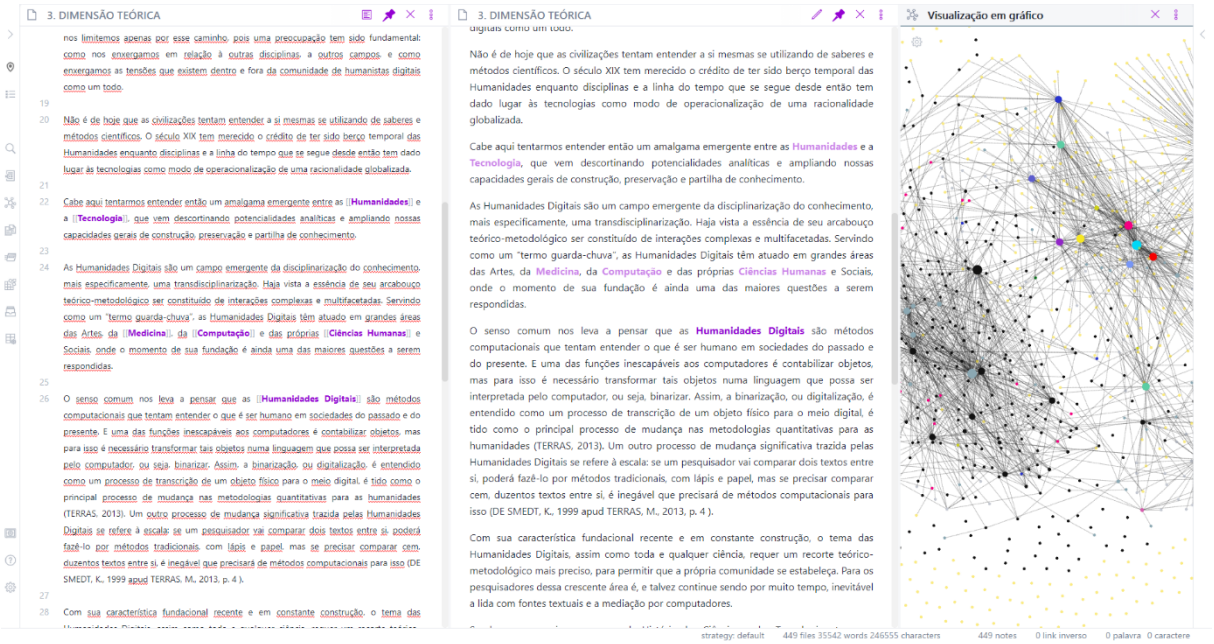
Fonte: *Obsidian*/Autor

Como parte de seu código interpretativo, se utiliza de uma linguagem computacional chamada *Markdown*, em que a utilização de alguns operadores linguísticos determina como a informação é apresentada na interface do *software*. Por exemplo, a utilização de colchetes duplos “[[ ]]” para um termo ou sentença determina a existência de uma referência dentro de uma dimensão textual, como também a referência imagética do grafo. Como parte da experimentação de uma metaescrita na pesquisa, utilizou-se o *Obsidian* para transgredir o processo de escrita tradicional, levando o conteúdo textual para a interface do *software* na intenção de representar possíveis relações entre as entidades e conceitos identificados.

Como vemos, a representação em grafo de uma escrita textual torna possível a releitura de narrativas e a produção de novos processos cognitivos, a partir de uma cartografia epistêmica, em que cada entidade, representada na imagem abaixo (Figura 2), é em si um elemento dotado de múltiplas relações. O grafo tem propriedades peculiares que fazem referência à Teoria de Redes, explorada mais a frente, em que podemos compreender a complexidade das relações estabelecidas entre as entidades, como o número de conexões, o peso de influência de cada

entidade, e a possibilidade de compreender os caminhos que tornam elementos distantes conectados entre si, ou seja, também faz referência ao conceito de *Seis Graus de Separação*[4].

Figura 2: Representação imagética em grafos de conceitos textuais formatados em Markdown

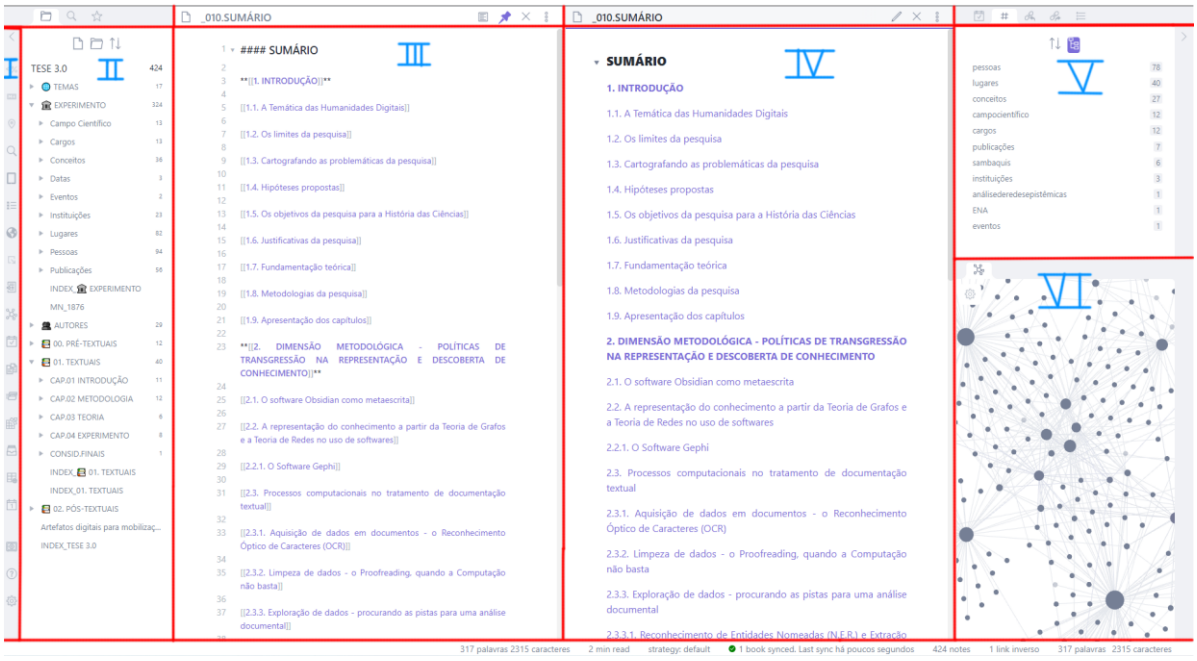


Fonte: *Obsidian*/Autor.

Ao propor a utilização de um *software* como o *Obsidian*, provocamos questionamentos sobre a estrutura linguística e cognitiva na formação e correlação de ideias percebidas sobre o mundo a partir de uma interface computacional com uma linguagem estruturada de modo que a mesma possa ser ampliada a partir de uma perspectiva "modular", ou seja, uma extensão de suas próprias capacidades de representação de informações, de possibilidades de descoberta de conhecimento. Em termos mais práticos, a interface do *Obsidian* que analisamos a seguir permite a constituição de processos cognitivos a serem definidos caso a caso na relação entre humano e máquina, por meio da manipulação de sua linguagem e seus códigos.

[4] Ver WATTS, 2003.

Figura 3: Interface do *Obsidian* apresentando seções estruturadas com funções distintas



Fonte: Obsidian/Autor

Como indica a imagem acima (Figura 3) em comparação a outra imagem (Figura 2), a interface pode ser dividida de maneira personalizada, obedecendo a critérios específicos seguindo uma lógica de visualização de informações de cada banco de informações elaborado. Assim, o *Obsidian* foi segmentado de modo que fosse possível a visualização do texto escrito tanto como os códigos (itens III e IV, em azul) implementados para a construção dos grafos, mostrados do lado inferior direito da imagem (item VI, em azul), recurso já mencionado anteriormente.

A interface também apresenta a possibilidade da listagem de *hashtags* (item V, em azul) relacionadas ao conjunto de arquivos presentes no que o *Obsidian* chama de ‘’cofre’’ (item II, em azul), que em termos práticos é uma pasta presente no armazenamento do computador. O uso de *hashtags* numa proposta transdisciplinar de metaescrita nos ajuda a criar pontes entre as entidades presentes em um texto, de modo que haja uma relação identificada visualmente no grafo.

O uso de *hashtags* faz parte de reflexões acerca do que entenderemos mais à frente como *pequeno-mundo*, e no campo da computação podemos analisar a

formação de *clusters* de entidades, agrupamentos que caracterizam entidades de valores similares, ou também podemos dizer que clusters são *comunidades epistêmicas*[5].

Seguindo com a descrição da interface do *Obsidian*, na lateral esquerda da interface se apresentam ícones (item I, em azul) com funções específicas, extensões que são adicionadas a partir de um repositório online que oferecem funcionalidades adicionais ao *Obsidian*, e que podem ser acionadas por meio de clique ou teclas de atalho. Uma dessas extensões possibilita o experimento do conceito de Seis Graus de Separação, mencionado anteriormente, em que a indicação de uma referência inicial e outra final pode revelar as mínimas conexões necessárias que as interligam, o que nos permite traçar novas narrativas entre entidades expressas no texto até então não percebidas claramente, pois são constituídas de modo linear.

2.2 A REPRESENTAÇÃO DO CONHECIMENTO A PARTIR DA TEORIA DE GRAFOS E DA TEORIA DE REDES COM USO DE *SOFTWARES*

A presente proposta teórico-metodológica desenvolve novos modos de ler e compreender a construção de conhecimento a partir de uma narrativa não-linear. Com o desenvolvimento de aplicações que se utilizam de algoritmos baseados na teoria de grafos, é possível criar visualizações computacionais-imagéticas de múltiplas relações entre entidades quando extraídas de documentação textual.

Tal processo cartográfico conta com uma abordagem transdisciplinar, haja vista a demanda por técnicas e ferramentas que pertencem a diversos campos do conhecimento, e em se tratando de uma pesquisa em História das Ciências, temos a oportunidade de explorar aqui esses atravessamentos enquanto diálogo entre a Teoria de Redes e a Representação e Descoberta de Conhecimento.

Representação do Conhecimento tem a ver com traduzir, transcrever, ou mesmo transgredir, em alguma linguagem ou meio de comunicação, descrições ou

[5] A clusterização de entidades aqui também é identificada em autores como Kuhn e Price, ao trazerem à luz das ciências as noções de *paradigma* e *redes científicas*, respectivamente. (KUHN, 1962; PRICE, 1965)

imagens que correspondem de alguma maneira o mundo percebido ou um estado parcial do mundo num tempo e espaço específico (BRACHMAN, 1985). Num contexto em que utilizamos *softwares* para realizar uma tarefa de representação, é primordial que suas interfaces e suas estruturas programáticas sejam capazes de interpretar os códigos inseridos de modo que possam chegar a novas conclusões, formalmente elaboradas, sobre essas descrições de mundo.

Tal capacidade interpretativa depende de uma "base de conhecimento", constituída previamente, para que sirva de referência simbólica sobre as características do que se quer interpretar de modo formal[6]. Uma das lacunas ainda problemáticas no que diz respeito ao uso de *software* para representação de conhecimento é a potencial incompatibilidade de se comunicar com plataformas que usam linguagens diferentes. Por isso a importância de se criar uma base de conhecimento que faça uso de códigos padronizados à prova de erros de transcrição entre uma interface e outra.

Além disso, é inegável perceber as dificuldades em representar e racionalizar a complexidade do mundo real numa interface computacional. Haverá sempre elementos derivados de outros tantos que não foram interpretados por completo; histórias que não foram passadas por meio de tradições orais, escritas em cartas ou registradas nas memórias das instituições.

No que diz respeito a utilização de ferramentas baseadas na teoria de grafos, o *Gephi*[7] é o *software* mais conhecido e bem explorado para a tarefa. Ainda que tenhamos uma formação tradicional na disciplina de História, ou em qualquer outra área tradicional das Humanidades, a experimentação com recursos computacionais

---

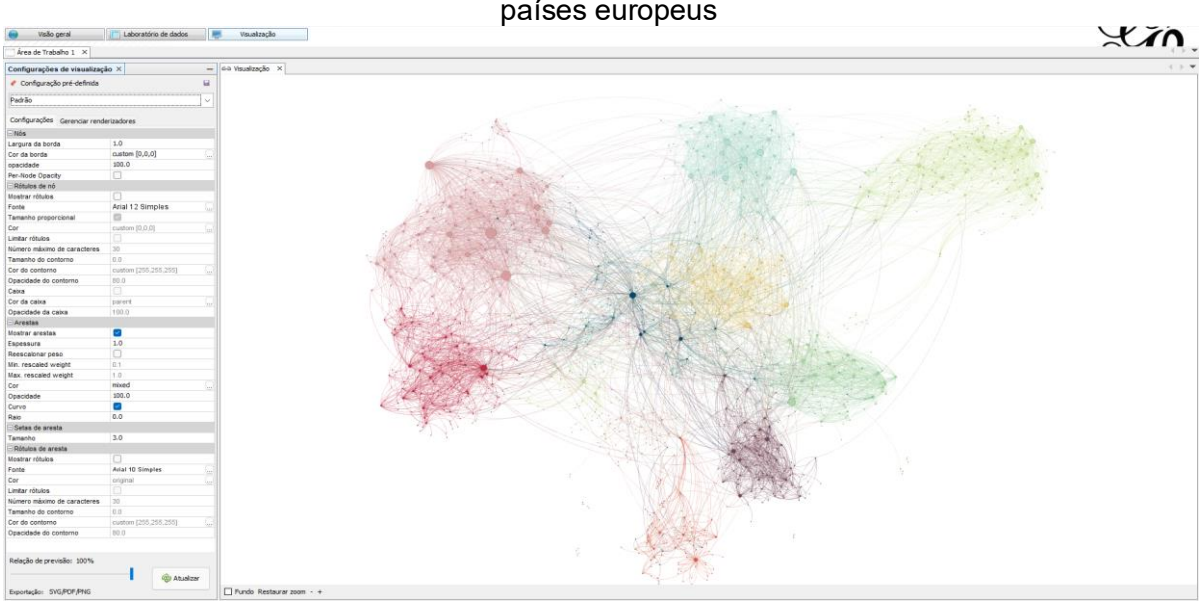
[6] Chegamos a um momento em que devemos nos perguntar como esse modelo de racionalidade, baseada em representações visuais, contribui com processos cognitivos e decisórios. Enquanto uma cartografia, no sentido deleuziano, processos de construção de redes requerem um foco direcionado naquilo que se apresenta na relação com o observador, como um território inexplorado. A primeira vista pode parecer uma aproximação com o positivismo lógico, haja vista a necessidade de uma consulta de verificação à uma base de conhecimento previamente constituída (e é aí que encontramos a contradição desse modelo racional, pois é impossível constituir um senso prévio de algo desconhecido), no entanto, a constituição das percepções de um mundo complexo se dá apenas com a inserção do observador no próprio território a ser explorado, em que este cria sua base de conhecimentos a partir dos elementos descobertos à medida que avança na relação do seu olhar para com o mundo. (POPPER, 1962; KUHN, 1962; DELEUZE, GUATTARI, 1980)

[7] CHERVEN, 2015.

como o *Gephi* pode proporcionar uma ampliação do olhar metodológico, teórico e experimental na constituição de nossos processos cognitivos e no desenvolvimento eficiente de um amplo olhar na lida com os objetos de pesquisa, especialmente as fontes textuais.

Diferente do *Obsidian*, que constitui visualizações em grafo por meio de identificadores da linguagem *markdown*, o *Gephi* tem a capacidade de importar informações em grande escala, de bancos de dados em planilhas, geralmente em formato *.csv, onde constam informações distintas, separadas entre as linhas e colunas. Assim, é possível selecionar uma gama de algoritmos capazes de construir visualizações diferentes, a depender da narrativa a ser dada ao conjunto de entidades representadas no grafo. Como exemplo para descrever algumas potencialidades do *Gephi* utilizou-se aqui um banco de dados público que mapeou as interações entre atores da sociedade civil e comunidade científica na internet no contexto de 12 países europeus.[8]

Figura 4: Mapeamento da interação entre sociedade e comunidade científica em países europeus



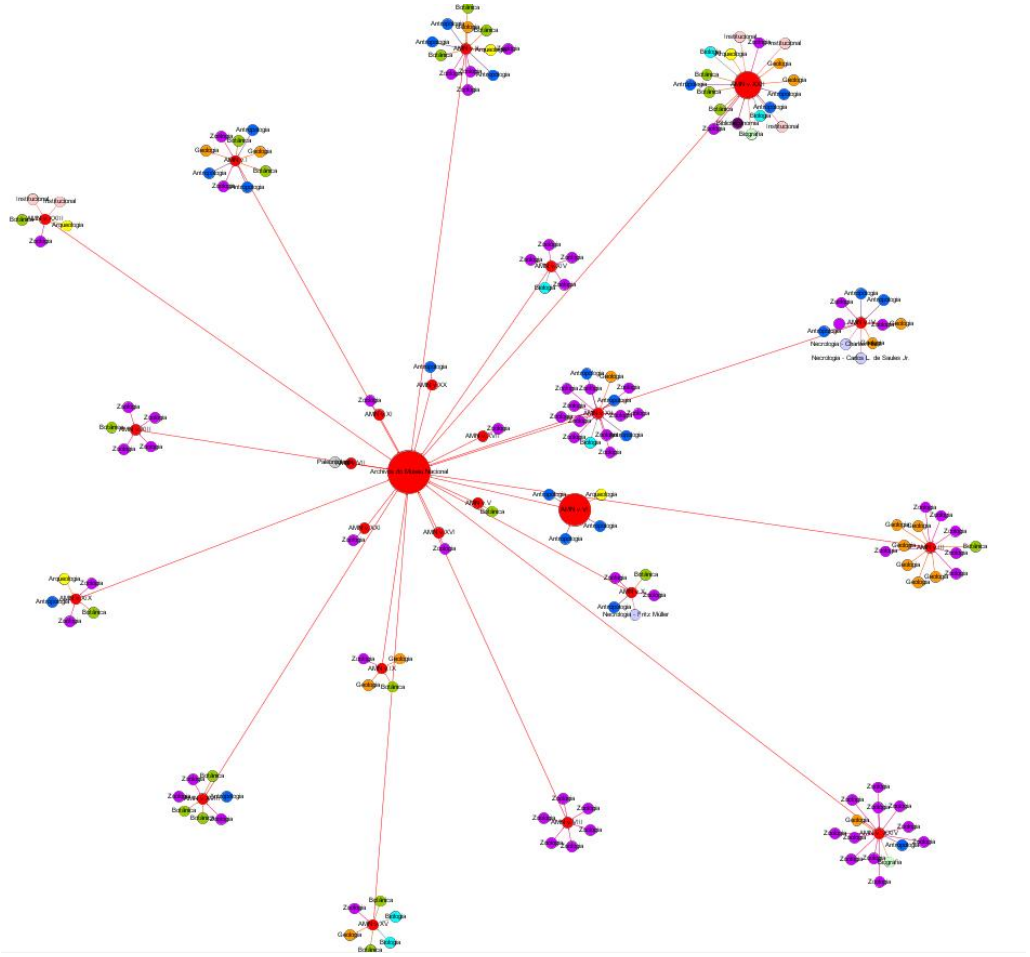
Fonte: Gephi/Autor

[8] **EuroSiS Web Mapping Study**. Disponível em: https://*Gephi*.org/datasets/eurosis.gexf.zip. Acesso em: 25/05/2021.

De maneira similar ao *Obsidian*, o *Gephi* também possui em sua interface um espaço para visualização completa dos grafos que interpreta. No entanto, a possibilidade de uso de algoritmos para manipulação das visualizações torna o *Gephi* um *software* muito mais complexo e com uma curva de aprendizagem bastante íngreme para pesquisadores que não tem formação ou familiaridade com essas metodologias.

Ainda assim, a partir de explorações teóricas sobre a teoria de grafos e a teoria de redes foi possível, mesmo manualmente, realizar experimentações com o *Gephi* em momentos anteriores no mestrado, com a documentação do Museu Nacional.

Figura 5: Representação das produções científicas do periódico Archivos do Museu Nacional



Fonte: Gephi/Autor

Ao utilizar o *Gephi* no exemplo acima (Figura 5), foi possível extrair informações textuais do periódico *Archivos do Museu Nacional* e transcrevê-las para um formato híbrido, codificadas como grafos. Percebe-se também que o uso de cores é um recurso incremental na teoria de grafos na constituição de agrupamentos, com o objetivo de marcar propriedades similares entre os demais nós da rede.

Tal recurso nos permite analisar também uma “capilaridade” entre os elementos representados de modo que suas características essenciais são fatores importantes para a constituição das conexões estabelecidas, sejam estas características parte de um campo epistêmico em comum ou um evento relevante para a compreensão dos processos que constituem uma rede.

2.3 PROCESSOS COMPUTACIONAIS NO TRATAMENTO DE DOCUMENTAÇÃO TEXTUAL

Uma cartografia que se desenvolve para cada conjunto de objetos, diferenciando-se a medida de suas complexidades transforma um texto sempre num novo experimento, resultando em novas fontes primárias. Os processos descritos aqui fazem parte de uma experimentação transdisciplinar, assimilando métodos, técnicas, ferramentas e conceitos da computação na aplicação de análise textual de documentação histórica. Etapas que consideram desde o estado de conservação do texto até suas potencialidades analíticas com a transgressão de seus signos, do analógico para o digital, de uma dimensão linear na leitura de um texto, para a apresentação de uma nova fonte, híbrida, imagética e não-linear.

2.3.1 *Aquisição de Dados em Documentos: o Reconhecimento Óptico de Caracteres (OCR)*

Descrever o processo de Aquisição de Dados textuais passa por uma experimentação condicionada pelas particularidades do documento processado. A complexidade dos ruídos de uma imagem determina os passos seguintes. Assim, a depender das condições físicas de preservação de um documento histórico e dos

recursos tipográficos empregados, as etapas de pré-processamento computacional podem ser mais ou menos complexas.

Devido o desenvolvimento do campo da computação e suas aplicações no processamento de documentos, o número de projetos de digitalização em instituições de memória tem aumentado exponencialmente nas últimas décadas. Com o intuito de acelerar a conversão da base documental para o formato digital, instituições precisam adotar metodologias eficientes para garantir a preservação da mídia física como também elaborar novos protocolos para as políticas científicas de base documental no formato digital.

Em se tratando de documentação histórica, o modo de preservação digital mais comum é o de cópias escaneadas, e de modo mais complexo, cópias que passam pelo Reconhecimento Óptico de Caracteres (OCR, em inglês). Sendo essa a etapa mais complexa na presente proposta teórico-metodológica transdisciplinar. Sem um documento completamente digital, capaz de ser lido por *softwares* e seus algoritmos específicos, dificilmente é possível dar seguimento a processos como a mineração de texto e a elaboração de grafos.

No entanto, ainda hoje as metodologias mais avançadas em OCR não são completamente eficazes para o processamento de documentação histórica, seja pela análise do *layout* dos documentos, por sua escrita tipográfica ou manual, ou mesmo o estado de preservação, como dito anteriormente.

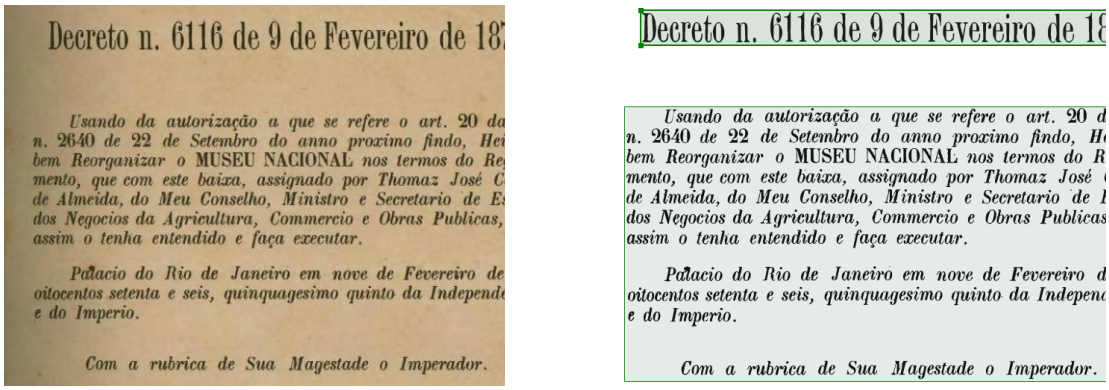
Como parte da experimentação documental, o processo de aquisição da documentação histórica analisada nesta pesquisa passa pela base de dados presente no acervo digital do Museu Nacional/UFRJ na seção da Comissão de Publicações, como também, em complemento, pela *Biodiversity Heritage Library* (BHL)[9].

Antes de realizar o OCR, etapas de pré-processamento são necessárias para aprimorar o reconhecimento dos caracteres, incluindo a aplicação de *threshold* ou limiar, ou seja, a identificação de valores máximos e mínimos entre o preto e o branco,

[9] Entre 2015 e 2021 verificou-se que o portal virtual do Museu Nacional passa por modificações significativas em sua interface, em que muitas vezes o acesso a documentação se encontrava indisponível. O portal BHL, por sua vez, conta com todo o acervo do periódico *Archivos do Museu Nacional* pesquisado para o período analisado, apresentando uma interface mais completa de recursos para recuperação de informações, inclusive com a transcrição do conteúdo do periódico para o formato digital (OCR).

para que a diferença entre caracteres e fundo se torne monocromática, facilitando a ação do algoritmo.

Figura 6: Aplicação de "threshold" como contraste entre preto e branco, um pré-processamento de imagem antes do OCR

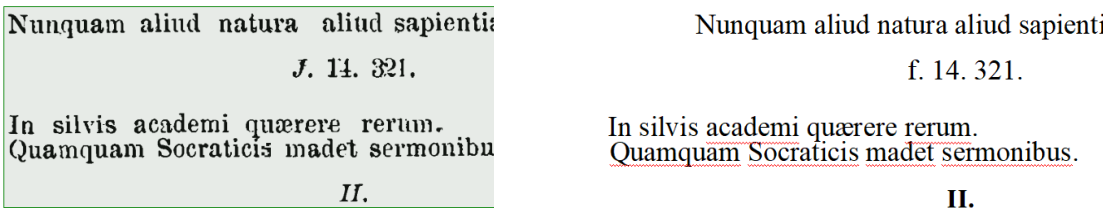


Fonte: Abbyy FineReader/Autor

Ao mesmo tempo que identificamos a necessidade do estabelecimento de contraste, também é necessário criar meios de diferenciar as seções de texto no documento analisado: fontes de tamanhos diferentes podem representar um título de uma seção ou mesmo a indicação de notas de rodapé, para tanto, uma análise de layout documental é de suma importância se quisermos preservar o formato do documento original, transcrito para um formato digital.

As análises tradicionais na utilização de OCR geralmente detectam e segmentam palavras e caracteres isolados nos documentos. No entanto, espaços entre caracteres não são sempre evidentes devido sua tipografia específica ou mesmo por deterioração da mídia analisada, o que certamente leva a erros de reconhecimento. Com o desenvolvimento de novos métodos em aprendizagem de máquina, o processo de reconhecimento tem se tornado cada vez mais eficaz. (KRÁL; LENC; MARTÍNEK, 2020)

Figura 7: Reconhecimento do caractere especial do latin "æ", a partir de padrão de reconhecimento interno do Abbyy FineReader 15.



Fonte: Abbyy FineReader/Autor.

Como vemos no exemplo acima (Tabela 1), a ligadura “æ” não oferece um espaço evidente entre os caracteres isolados, tornando o processo de OCR dependente de aprendizagem complementar, se de modo manual, pela inserção de um novo dado pelo usuário, ou de modo automático, por meio de um banco linguístico previamente constituído.

2.3.2 *Limpeza de dados: o Proofreading, quando a computação não basta*

Em se tratando de um processo de normalização, a limpeza de dados é uma etapa que segue as diretrizes formuladas em etapas anteriores, respeitando objetivos bem definidos, sejam eles de restauração e manutenção do documento original ou transgressão da informação para novos meios.

Dois fenômenos podem se apresentar após a realização do processo de OCR: primeiramente, o reconhecimento parcial dos caracteres, incorrendo em erro de grafia; segundo, e como consequência do primeiro, a ausência de um parâmetro comparativo com uma base de palavras ou dicionário base, incorrendo em sinalização de erro em *softwares* processadores de texto, como o *Word* ou *LibreOffice*, ou mesmo processadores de texto internos, como no caso do *Abbyy FineReader*.

Figura 8: Processamento parcial de caracteres resultando em erros de grafia e ausência de comparativo com dicionário interno, devido diferenças com o português arcaico.

Fonte: Abbyy FineReader/Autor

Em ambos os casos é necessária uma intervenção humana no processamento dos documentos. Até o momento não há solução automatizada para a etapa de correção da grafia dos caracteres, apenas sugestões baseadas em aprendizagem de máquina, e mesmo assim, dependente de decisão humana.

Uma base de dados linguística é necessária para estabelecer padrões de reconhecimento de caracteres em que algoritmos de inferência possam sugerir correções ao processo de OCR. Utilizar dicionários de idiomas específicos, como também dicionários constituídos pelo usuário é uma maneira de tornar mais preciso todo o processo de correção, ainda que semiautomático[10].

[10] No exemplo acima, enquanto estudo de caso do periódico *Archivos do Museu Nacional*, explorado mais à frente, foi necessário a adição de termos do português arcaico, haja vista ser uma documentação do século XIX. Tal processo demanda tempo considerável do pesquisador, o que retoma a discussão sobre a falta dicotomia entre eficiência e automação da pesquisa no campo das humanidades digitais. Ou seja, a utilização de métodos computacionais, além de potencializarem, e de certo modo acelerarem, os processos de análise documental, também fazem emergir novos problemas e novas demandas de aprendizagem por parte de pesquisadores das humanidades.

Em experiência para tentar solucionar essa questão, verificou-se a existência de serviços pagos online de correção, o que não atendia aos parâmetros da pesquisa[11]. O mais próximo que identificamos como um processo de correção colaborativo foi o portal *WikiSource*[12], parte da *WikiMedia Foundation*, em que encontramos exemplares de documentação histórica transcritas e preservadas por um coletivo de usuários que inserem os dados manualmente (ANEXO 1).

2.3.3 *Exploração de dados: procurando as pistas para análise documental. Reconhecimento De Entidades Nomeadas (N.E.R.) e Extração de Entidades Nomeadas*

A Exploração dos Dados é também um processo cartográfico na construção de um banco de dados de entidades nomeadas. A construção de um modelo analítico posterior vai depender da exploração do documento “em tempo real”; somente com a identificação das entidades é que poderemos constituir um banco de dados eficiente e traçar narrativas com novos olhares, de um texto com leitura linear, para um grafo de leitura não-linear.

N.E.R trata de métodos de identificação de entidades como pessoas, locais, datas, conceitos, localizados em documentação textual variada. Como parte integrante do campo de Processamento de Linguagem Natural (NLP, em inglês), o reconhecimento de entidades serve de processo de automação, dependente de processo anterior como o OCR. O que estabelece de certo modo uma cadeia de processos computacionais que transformam o objeto físico e seu conteúdo, em objeto digital.

Transferindo também suas características morfológicas, o reconhecimento de entidades nomeadas é passível de imprecisões caso o OCR realizado no texto

---

[11] A intensão da pesquisa, por mais que tenhamos utilizado softwares proprietários como experimento, é a de promover, enquanto política científica, o uso de software livre e gratuito, num esforço de colaboração e padronização institucional.

[12] **WikiSource**. Disponível em: https://pt.wikisource.org/wiki/Wikisource:página_principal. Acesso em: 25/05/2021.

confunda, por exemplo, a capitalização dos tipos de caracteres, tornando nomes próprios em termos comuns; variações no léxico também podem tornar a tarefa da identificação de entidades nomeadas um processo difícil de ser realizado sem correções mesmo nos algoritmos utilizados (SILVA, 2020). A fim de evitar essas imprecisões, o reconhecimento de entidades nomeadas também depende de um banco de dados linguístico específico para cada idioma e temática analisada.

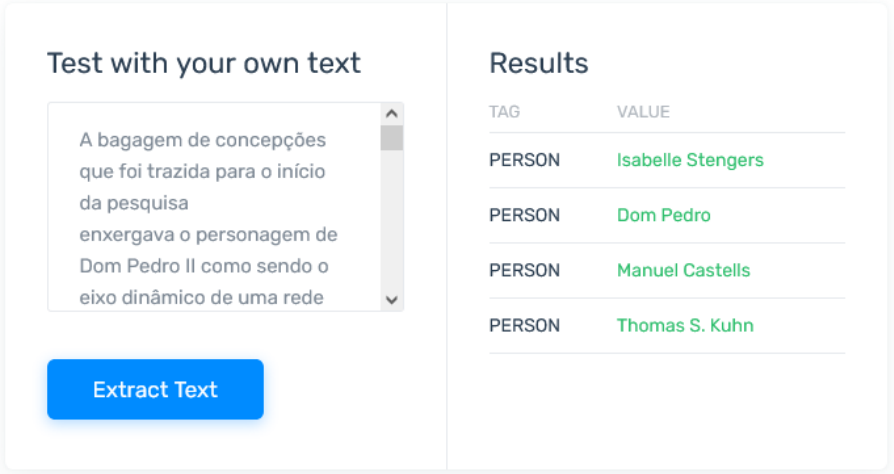
No estudo de caso desta pesquisa, esse banco de dados linguístico consiste na identificação de sujeitos, lugares e instituições que pertencem ao espaço de fluxo de produção de ciências no contexto brasileiro, como também referências a sujeitos que transitam em espaços externos, como instituições nacionais e internacionais, congressos e exposições universais (ARCHIVOS DO MUSEU NACIONAL, 1876).

No exemplo a seguir, verificamos como o modelo de N.E.R extrai a informação relevante a partir de um contexto em que reconhece características em comum entre as entidades, classificando-as e listando-as no resultado da análise. Assim, aplicando o método N.E.R em um parágrafo extraído de texto referente a pesquisa de mestrado desse autor, temos:

> A bagagem de concepções que foi trazida para o início da pesquisa enxergava o personagem de Dom Pedro II como sendo o eixo dinâmico de uma rede egocentrada, uma qualidade psicanalítica centrada no indivíduo e em ações subjetivas na constituição dos agenciamentos da rede. Com o objetivo de constituir as bases teórico-metodológicas da pesquisa, um conjunto de conceitos presentes na área da sociologia, tendo como expoente Manuel Castells e o espaço de fluxos, da ciência da informação, com a bibliometria e a formação das comunidades científicas a partir da obra de Derek J. De Solla Price, e as questões paradigmáticas e dos limites científicos expostos por Thomas S. Kuhn e Isabelle Stengers, foi possível observar um traço comum: todas essas abordagens teóricas tem as redes como uma realidade abstrata na construção de suas análises. (MAIA, 2017).

Em seguida, o algoritmo responsável pela identificação de entidades nomeadas lista os resultados com a categoria encontrada e todas as entidades pertencentes a ela.

Figura 9: Identificação de Entidades Nomeadas na categoria "Pessoas".



Fonte: MonkeyLearn/Autor

Notemos que, mesmo com um processo automatizado e um algoritmo especializado em extração de informações que tem uma base linguística para comparações, alguns nomes não foram reconhecidos e outros apenas parcialmente. Supomos que a explicação para esse resultado aproximado, se deve a um algoritmo mal programado ou uma base de dados linguística limitada.

2.3.4 *Modelagem de dados: reconhecimento de padrões e replicabilidade analítica*

A temática da modelagem de dados é cara tanto para a historiografia tradicional como para a computação e suas metodologias assimiladas cada vez mais pelas humanidades digitais. Numa brevíssima retrospectiva metodológica, lembremos da importante obra Domínios da História (CARDOSO & VAINFAS, 1997), em que se investiga a diversidade dos territórios que historiadores habitam desde Heródoto para a formalização de nossa atividade intelectual e disciplinar.

Desde uma história política, econômica e social, passando pela história das ideias, das mentalidades e da diversidade cultural, até chegarmos à emergência dialógica da microhistória, nos trazendo até aqui, com movimentos transgressores e transdisciplinares, vemos cada vez mais a adoção de modelos metodológicos para as

interpretações, análises e mesmo previsões históricas com o uso de dados seriados e informações estatísticas com o auxílio de banco de dados e ferramental computacional.

Com a multitude de abordagens metodológicas, o advento da computação para desenvolvimento e mediação na descoberta e visualização de conhecimento, a demanda por organização e modelagem de dados (FLANDERS & JANNIDIS, 2019) tem sido cada vez mais presente nas discussões nas humanidades digitais como um tópico crítico enquanto análise da dimensão semântica dos sujeitos e objetos representados e processados pela interação humano-máquina, como por exemplo, em visualizações em grafo, método adotado nesta pesquisa.

Como parte de um esforço colaborativo nas discussões teóricas sobre as práticas em humanidades digitais, Julia Flanders e Fotis Jannidis exploram essa dimensão organizacional e modelagem de dados no contexto das humanidades, valendo como diálogo essencial no estabelecimento desta etapa no que diz respeito aos tipos de processos computacionais para o tratamento de documentação textual histórica.

Em que pese as constantes questões formuladas sobre os pilares teóricos das humanidades digitais, ainda nos parece inescapável a adoção de abstrações tradicionais das humanidades na definição da práxis nesse novo campo com o auxílio de processos computacionais.

Trazendo os objetos analisados pela computação como condicionantes nas análises feitas em humanidades digitais, os bancos de dados são interpretados como objetos que constituem estruturas fundamentais nos sistemas de informação, em que cada conjunto de dados deve ser interpretado de modo situado, ainda que contenham características que permitam sua aplicação em objetos similares, por exemplo, em documentação seriada de instituições.

Uma “modelagem de dados”, no campo da ciência da computação consiste em compreender uma série de etapas no processamento de informações em interação com seus operadores, que juntos constituem um mecanismo abstrato, sendo a modelagem o processo que leva a visualização e, por fim, a análise crítica das

informações enquanto retorno da interação humano-máquina (FLANDERS & JANNIDIS, 2015, p. 22).

Uma primeira etapa, conceitual, para identificar e descrever as entidades presentes numa determinada dimensão textual, por exemplo, como também as relações estabelecidas entre si; uma segunda etapa, lógica, para definir a qualidade dos classificadores de entidades, separando-os semanticamente numa tabela que possa ser lida pelo computador; uma terceira etapa, física, buscando otimização para implementação do modelo de dados.

No entanto, é na constituição lógica e semântica do modelo de dados que posteriormente podemos definir, explicar e replicar o modelo de dados na construção de uma topologia de rede (grafos), ao analisarmos entidades a partir de um conjunto documental enquanto fonte histórica primária.

Para tanto, o processo de modelagem de dados deve ter como princípio uma curadoria e pesquisa relacional das fontes a serem extraídas direcionadas pelo contexto do objeto analisado, como um sistema de controles, ou seja, é preciso situar a fonte analisada definindo os limites de aplicabilidade para o modelo a ser constituído com suas entidades. Dados extraídos de um conjunto de documentos textuais históricos que traduzem uma dimensão histórica, social, econômica etc., não devem ser usados para criar um modelo que representem dimensões estranhas ao contexto das fontes.

Por fim, o modelo como uma estrutura instanciada, um produto com arestas definidas, com diretrizes próprias para aplicação em análises específicas. Um processo de suma importância para o campo das humanidades digitais por prover artefatos de pesquisa com os quais novos métodos e ferramentas podem ser desenvolvidos e aplicados.

Tais estruturas podem vir representadas por um conjunto de diretrizes linguísticas, como uma lista de palavras extraídas de um determinado conjunto documental, o que caracterizaria um banco de dados linguístico ou um dicionário personalizado, contendo expressões e referências espaço-temporais específicas; estruturas instanciadas que contêm comandos de execução de tarefas se utilizando de linguagens de programação de computador em interação com os conjuntos

documentais disponíveis, como a linguagem *python*, muito utilizada em tarefas baseadas em aprendizagem de máquina, inteligência artificial e mineração de dados.

2.4 POR UMA METODOLOGIA CRÍTICA EM HUMANIDADES DIGITAIS

Pensar uma nova estatística baseada na interação humano-máquina sendo mediada por interfaces que permitem uma maior gestão da informação científica e histórica é uma proposta crítica num momento de emergência de narrativas que bombardeiam o fazer e pensar científico com uma infinidade de dados, verdadeiros e falsos, estes produzidos não só por sujeitos humanos, mas também por máquinas e organismos, capazes de traduzirem suas dimensões existenciais para a compreensão humana.

Procedimentos estatísticos que têm dado diferentes nomes às novas práticas contábeis do conhecimento, sejam essas práticas compreendidas como um formalismo quantitativo, leitura distante, critica algorítmica, macroanálise ou mineração de dados, cada uma dessas denominações realiza, de uma forma ou de outra, alguma forma de computação. E a premissa na adoção de métodos, técnicas e ferramental sofisticados é objetivo central dos humanistas contemporâneos para a exploração, investigação e preservação dessa miríade de dados que permanecem “codificados” nos textos históricos e científicos, em sua maioria ainda não processados com a ajuda de computadores.

Esse contexto de transição de objetos físicos para digitais faz parte do que Jeffrey Schnapp e Todd Presner estabeleceram no seminal “Manifesto das Humanidades Digitais”[13] como a “primeira onda” de transformações epistemológicas na institucionalização do campo (DOBSON, 2019, p. 3). A segunda onda de transformações se caracteriza pela mudança para estudos de novas mídias e outras fontes nativas digitais em que não requerem conversões em sua natureza.

Uma terceira onda de transformações epistemológicas do campo das humanidades digitais estaria ocorrendo no momento presente, em que o principal

[13] ANEXO 2.

objeto de pesquisa seria a noção de “computacionalidade”, proposta por David M. Berry em que há de fato uma mudança ontológica sobre a compreensão histórica dos sujeitos, artefatos, fluxos e dimensões do digital (BERRY, 2011, p. 16).

Berry chama atenção sobre os *softwares* e *hardwares* que tornam possível o processamento dos objetos dos contextos epistêmicos anteriores, culminando também no desenvolvimento e ampliação dos chamados *software studies* como parte desse potencial analítico, e que para entender tais operações é preciso considerar criticamente o território habitado pelos sujeitos das humanidades digitais para que se analise essa computacionalidade enquanto tal ao mesmo tempo que trazem à tona questões que apresentam essa terceira onda como um “deslocamento ontológico” (DOBSON, op.cit).

Uma crítica também levantada por Katherine Hayles delineia duas possíveis práticas operativas das humanidades digitais em relação a outros campos científicos, como a Biologia, Psicologia, Antropologia e a Linguística: assimilação e distinção. A primeira parece estender a prática de pesquisa acadêmica para a dimensão digital, oferecendo mais acessibilidade que os formatos impressos, principalmente nas capacidades de busca e recuperação de informações e na publicização destas, uma prática mais garantista do que de confronto epistêmico propriamente dito, apenas o incremento na adoção de novas ferramentas digitais.

A segunda enfatiza novas metodologias, o surgimento de novos questionamentos e a emergência de novos campos do conhecimento. Num movimento contrário, a prática da distinção faz parecer as humanidades digitais menos Humanidades ao se apropriarem de métodos e questões dos outros campos, deixando o hábito do criticismo no passado ao confrontar e tornar obsoletos os métodos e questões de campos do conhecimento das humanidades (HAYLES, 2012).

Como um apontamento particular enquanto formulação dessa dimensão metodológica da presente pesquisa, e levando em consideração as práticas locais no Brasil sobre o desenvolvimento das humanidades digitais, não é útil nem socialmente desejoso desconsiderar um avanço em direção a um olhar transdisciplinar, onde seja possível não só a assimilação de métodos e técnicas de campos distintos das

humanidades, como também o desenvolvimento de uma epistemologia crítica[14] sobre a natureza própria dos sujeitos e artefatos que porventura venham a adentrar as humanidades por meio de processos computacionais.

3 DIMENSÃO TEÓRICA: A EMERGÊNCIA TEÓRICO-METODOLÓGICA PARA A HISTÓRIA DAS CIÊNCIAS

Talvez a escolha do título deste capítulo faça justiça a investigação que ora se inicia. O significado da "emergência" das Humanidades Digitais, aqui, trata tanto de uma discussão epistemológica como da política desse campo. Primeiro, é compreensível que um novo campo e seus proponentes se preocupem com processos epistemológicos definidores de seus conceitos particulares; segundo, ainda que seja um campo sem uma definição cristalizada, alimenta, desde sua gênese, questionamentos sobre a construção de saberes por meio de seus múltiplos métodos e aplicações no seio das diferentes realidades sociais, políticas, econômicas e filosóficas que atravessam o seu próprio desenvolvimento. Sendo esta última o catalisador que traz à luz os desdobramentos de uma dimensão teórica.

Pelo menos uma dezena de perguntas vem à tona antes mesmo de começarmos a compreender a gênese de um novo campo do saber. Perguntas que requerem todo um cuidado e uma atenção epistemológica, até mesmo uma cartografia de suas enunciações. Ainda que precisemos no momento, por motivos óbvios e úteis, questionar as definições do campo das humanidades digitais, não nos limitemos apenas por esse caminho, pois uma preocupação tem sido fundamental: como enxergamos as tensões que se constituem dentro e fora da comunidade de humanistas digitais na apropriação, assimilação e transformação de métodos e técnicas de outros campos do conhecimento?

[14] O desenvolvimento de uma epistemologia crítica é parte do processo de emergência e imediatismo da absorção de múltiplos métodos e teorias, tanto das Humanidades como das Tecnologias; é preocupação fundante das Humanidades Digitais para o estabelecimento de um novo olhar transdisciplinar na imersão de uma complexidade contemporânea da realidade que se manifesta, e regressa, do Digital.

Não é de hoje que as civilizações tentam entender a si mesmas se utilizando de saberes e métodos científicos. O século XIX tem merecido o crédito de ter sido berço temporal das Humanidades enquanto disciplinas e a linha do tempo que se segue desde então tem dado lugar às tecnologias como modo de operacionalização de uma racionalidade globalizada.

Cabe aqui tentarmos entender então um amalgama emergente entre as Humanidades e a Tecnologia, que vem descortinando potencialidades analíticas e ampliando nossas capacidades gerais de construção, preservação e partilha de conhecimento.

As Humanidades Digitais são um campo emergente da disciplinarização do conhecimento, mais especificamente, uma transdisciplinarização. Haja vista a essência de seu arcabouço teórico-metodológico ser constituído de interações complexas e multifacetadas. Servindo como um “termo guarda-chuva”, em que acolhe teorias, métodos e técnicas de campos tradicionais, as Humanidades Digitais têm atuado em grandes áreas das Artes, da Medicina, da Computação e das próprias Ciências Humanas e Sociais, onde o momento de sua fundação é ainda uma das maiores questões que embasam o reconhecimento de um momento de virada epistemológica, de caráter transdisciplinar.

O imaginário sociotécnico tem constituído as Humanidades Digitais como métodos computacionais que tentam entender a relação dos humanos com artefatos e técnicas que auxiliaram as sociedades a se desenvolverem e a criar uma perspectiva de futuro por meio das ciências e tecnologias. E uma das funções inescapáveis aos computadores é contabilizar a realidade, mas para isso é necessário transformar tal realidade numa linguagem que possa ser interpretada pelo computador, ou seja, binarizar.

Assim, a binarização, ou digitalização, é entendido como um processo de transcrição de um objeto físico para o meio digital, tido como o principal processo de mudança nas metodologias quantitativas para as humanidades (TERRAS, 2013). Um outro processo de mudança significativa trazida pelas Humanidades Digitais se refere à escala: se um pesquisador vai comparar dois textos entre si, poderá fazê-lo por métodos tradicionais, com lápis e papel, mas se precisar comparar cem, duzentos

textos entre si, é inegável que precisará de métodos computacionais para isso (DE SMEDT, 1999 *apud* TERRAS, 2013, p. 4).

Com sua característica fundacional recente e em constante construção, o tema das Humanidades Digitais, assim como toda e qualquer ciência, requer um recorte teórico-metodológico específico, para permitir que a própria comunidade se estabeleça. Para os pesquisadores dessa crescente área é, e talvez continue sendo por muito tempo, inevitável a lida com fontes textuais e a mediação por computadores.

Sendo uma pesquisa no campo da História das Ciências e das Tecnologias, temos a oportunidade de avaliar fontes históricas textuais a partir da utilização de ferramentas computacionais, o que nos leva a construção de fontes híbridas, binarizadas, mas ainda textuais, de tal modo que nos permita uma proposição teórico-metodológica transdisciplinar como política científica de análise documental para a História das Ciências no Brasil.

Os processos de identificação das questões referentes à pesquisa dão continuidade a um processo anterior, de mesmo tema e abordagem. Enquanto pesquisadores da área das Ciências Humanas, mais especificamente a História, como podemos nos utilizar de métodos quali-quantitativos de análise documental mediados por computador? A lacuna na formação dos humanistas é uma questão séria a ser debatida aqui enquanto proposta teórica e de política científica, pois vai além de diretrizes pedagógicas.

Estamos chamando a atenção para uma linguagem convergente, dinâmica e que atravessa os diversos e, muitas vezes opostos, campos do conhecimento científico, como também para métodos de análise que identifiquem a complexidade das esferas social e política na construção do conhecimento histórico e científico e a percepção que a sociedade tem destes.

Assim, a questão que move a pesquisa é compreender como a adoção de múltiplos métodos de análise documental pode revelar novos caminhos na construção do conhecimento histórico e científico no contexto brasileiro. É possível para historiadores das ciências, pela mediação da tecnologia dos computadores, um novo fazer historiográfico? É com o reconhecimento da metáfora das redes que

construímos aqui os epicentros teóricos para uma abordagem transdisciplinar convergente sob o termo das Humanidades Digitais.

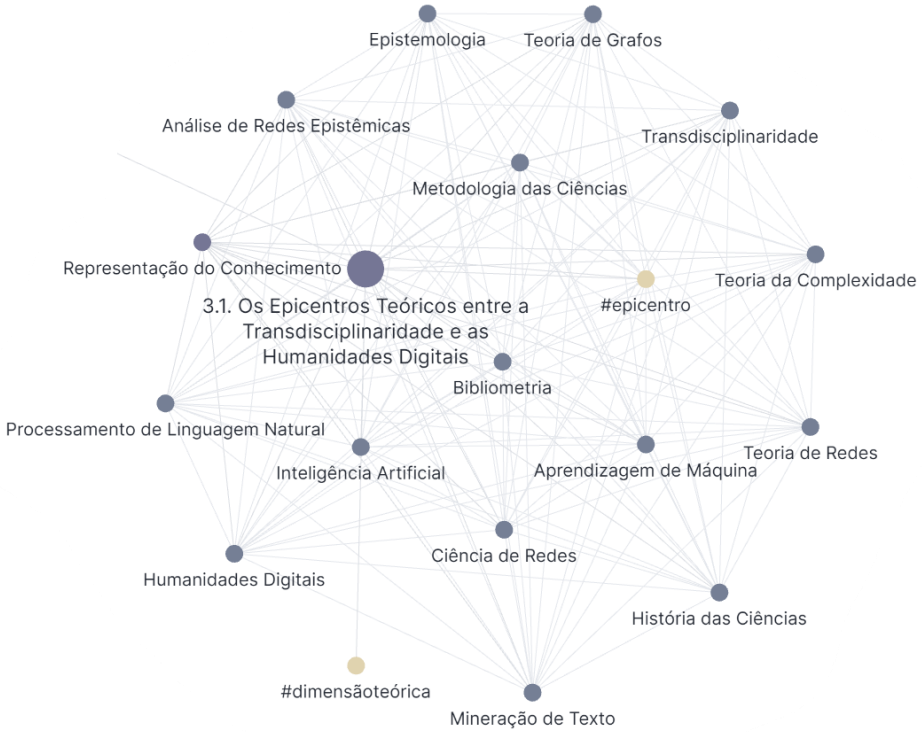
3.1 OS EPICENTROS TEÓRICOS TRANSDISCIPLINARES PARA AS HUMANIDADES DIGITAIS

A peculiaridade do campo das humanidades digitais requer um processo de investigação epistemológica que, além de passar pela identificação de seus sujeitos e os paradigmas que atravessam esse espaço de construção de conhecimento, também necessita de metodologias e teorias transdisciplinares à medida que as interlocuções entre disciplinas tradicionais vão se tornando cada vez mais complexas e codependentes.

De modo que haja um diálogo mais apropriado para a presente pesquisa, entre a metodologia desenvolvida e a elaboração do que chamamos nesse momento de “epicentros teóricos”, adotamos uma abordagem deleuziana que tem como metáfora o rizoma, uma cartografia processual em que cada sujeito e cada conceito nos fornecem pistas cognitivas para a construção, ou mesmo descoberta, de outros tantos sujeitos e outros conceitos relacionados aqueles primeiros.

Resolvemos assim, representar essa cartografia por meio de grafos (Figura 7), uma releitura dialógica que permite a visualização das interações de modo amplo na construção de possíveis narrativas não-lineares, um modo de representar e descobrir conhecimento que difere de uma leitura linear tradicional.

Figura 10: Representação de temas e campos do conhecimento cartografados a partir de revisão bibliográfica.



Fonte: Obsidian/Autor

Fazendo uso do *software* ***Obsidian*** é possível criar uma representação imagética interativa dos campos do conhecimento que fazem parte dessa cartografia processual que tem como metáfora as relações em rede. Partindo de uma alusão à ideia deleuziana de rizoma (DELEUZE, 1995), representações e descoberta de conhecimento se dão a partir de qualquer ponto de partida, considerando as correlações temáticas que se avizinham e complementam o entendimento do quadro geral observado.

Como começamos esse capítulo mencionando a metáfora das redes, cabe aqui uma exploração dialógica sobre os principais conceitos e perspectivas da Teoria de Redes, desde aspectos matemáticos, sociológicos até os aspectos computacionais em busca de pistas da complexa relação com os outros pilares teóricos. Assim, foi com Manuel Castells e sua investigação sobre a *Sociedade em Rede* (CASTELLS, 2005) que nos deparamos com o conceito de “espaço de fluxos”, uma dimensão

comunicacional que se constitui de múltiplas trocas entre sujeitos, ideias, técnicas, artefatos e políticas que, em conjunto, refletem contextos históricos sobre relações econômicas, sociais, como também científicas. É a partir dessa perspectiva sociotécnica da Teoria de Redes que a consideramos um dos pilares para uma epistemologia das Humanidades Digitais.

Ainda tomando como exemplo a representação temática na imagem anterior, é com uma abordagem transdisciplinar, de múltiplas conexões, que podemos constituir narrativas[15] que relacionam diversos campos como a “Representação de Conhecimento” e a “Teoria de Grafos”, intimamente ligada também à teoria de redes. A identificação dessas relações dá forma à rede, constituindo sua topologia, uma estrutura que possibilita analisar a força e a dinâmica das conexões entre os “nós”, constituindo também agrupamentos de entidades que tem propriedades específicas.

A “clusterização” de entidades de uma rede permite compreender a formação, por exemplo, de narrativas localizadas, quando se estabelecem similaridades, ou mesmo disparidades, entre os nós da rede[16].

Para Barabási, essas narrativas localizadas, resultado de fragmentações e ramificações de narrativas maiores e anteriores, parte de um processo de especialização na interpretação da realidade cada vez mais complexa, em consequência da edificação das chamadas ciências modernas, e no contexto do século XXI, com o desenvolvimento de olhares transdisciplinares, ainda não sabemos como remontar essas peças novamente (BARABÁSI, 2002. p. 6). Tomando a liberdade de parafraseá-lo, “quando pilotamos a locomotiva do reducionismo, damos de cara com a muralha da complexidade”.

A constituição de agrupamentos de entidades por meio de microrrelações também está presente no campo da “bibliometria” que, enquanto método de análise

---

[15] Mais adiante veremos que o estudo de caso realizado no contexto do periódico *Archivos do Museu Nacional*, o espaço de fluxos das ciências no século XIX passou por transformações e expansões importantes e que, à medida que identificamos os sujeitos, instituições, as ideias e as políticas relacionadas às ciências, foi possível compreender melhor as conexões para a elaboração de um mapeamento e uma representação computacional-imagética destas.

[16] Podemos interpretar a clusterização de entidades como a formação de “pequenos-mundos”, um conceito da teoria de rede que representa um olhar microscópico, ou de microrrelações, em comparação com o quadro geral analisado. Ver WATTS, 1999.

estatística para compreender a dinâmica da produção de conhecimento científico, é também atravessada pela metáfora das redes. Estatísticas que procuram descortinar estruturas e padrões de emergência epistêmica que revelam crises paradigmáticas kuhnianas ou padrões de expansão e rupturas de comunidades científicas tão bem investigados por Price[17].

Mais adiante, ao experimentar uma metodologia transdisciplinar de análise computacional-imagética, veremos que é possível não só reconstituir as narrativas transcritas em um documento histórico, mas também derivá-las para novos caminhos, ou mesmo caminhos inexplorados, muitas vezes silenciados devido a visões lineares na construção dos registros históricos. É preciso reconhecer aqui que uma dimensão linear do registro textual histórico e científico nos trouxe até o contexto atual do conhecimento humano, o que não impede de explorarmos os múltiplos caminhos e dimensões que por ventura se escondem nas entrelinhas da história escrita.

E apenas agora, com o aumento exponencial da capacidade tecnológica computacional, podemos recuperar ou mesmo constituir novos sentidos para resolver antigos problemas.

É a partir desse ponto que precisamos dialogar com disciplinas e saberes que se mantiveram alheios ao olhar dos humanistas até pouco tempo. Dois desses campos são o “Reconhecimento Óptico de Caracteres” (OCR) e a “Mineração de Texto”, metodologias derivadas da ciência da computação e da mineração de dados que também compõem os saberes relacionados a “Aprendizagem de Máquina” (*Machine Learning*, em inglês).

O OCR é uma técnica de recuperação de informações que tem sido aplicada há mais de quarenta anos em analise textual, tanto na indústria como na comunidade acadêmica, a fim de sanar a demanda da transição informacional digital devido ao grande volume de impressos e manuscritos da história humana. Textos que estão

[17] Cada um à sua maneira, Thomas Kuhn e Derek Price dialogam sobre a dinâmica de produção e desenvolvimento de conhecimento científico, se utilizando de métodos em campos científicos distintos, mas que revelaram demarcações similares: verdades científicas são finitas, constituídas sob um consenso de mentes humanas. Ler mais em *A Estrutura das Revoluções Científicas* (KUHN, 1962) e *Little Science, Big Science...and Beyond* (PRICE, 1963).

organizados em bibliotecas, em acervos pessoais ou mesmo registros de organizações e instituições públicas e privadas de todos os ramos.

O processamento de informações textuais em meios analógicos consome ainda hoje grande parcela de tempo de trabalho, por isso o OCR se tornou etapa essencial na transição para o digital. Ainda que este não ofereça uma eficácia de 100%, já avançou muito na qualidade do processamento de tipografias mecânicas e na construção de bancos de dados de manuscritos em múltiplos idiomas[18]. Um banco de dados constituído a partir da análise de pixels que caracterizam símbolos alfanuméricos ou mesmo de caracteres especiais, geralmente apresentados em contraste, em níveis de cinza, com o fundo de uma página.

Assim, os caracteres são binarizados e tendo seu formato reconhecido total ou parcialmente pelo computador. Sendo uma etapa indispensável da análise documental, o OCR representa um novo esforço metodológico e experimental para os humanistas, estes tendo que lidar com novas abordagens e fronteiras epistemológicas, introduzindo novos conceitos e novas visões a seus objetos de pesquisa.

A clusterização por sua vez, parte da metodologia de mineração de texto, é uma técnica de organização de informações relacionais em larga escala, que auxilia na localização de dados relevantes em uma determinada pesquisa, ou mesmo a identificação de ramificações de narrativas dentro de uma comunidade epistêmica. Geralmente se valendo de conceitos, identificadores de temas, ideias, pessoas, lugares etc, aqui chamados de entidades, a clusterização representa um agrupamento ou mesmo um sumário de um conjunto de documentos, postos sob a ótica da mineração de dados textuais, com objetivo de constituir análises estatísticas e descobrir potenciais novos padrões informacionais ou de novos dados sobre um determinado assunto (PRADO, 2007, p. 224).

Estando cercados de registros históricos documentais que abrangem diversos temas, a automação pelo processamento e digitalização requer cada vez mais técnicas e recursos computacionais capazes de lidar com tal demanda. Documentos

[18] CHERIET, 2007, p. 6.

que se encontram em diversos formatos e estados de conservação são agora transcritos, por meio de escaneamento, para um ambiente digital onde sua identificação e tratamento da informação são melhor processadas. Sabendo que o processamento manual de registros históricos sempre foi uma tarefa de extremo custo laboral, não só para historiadores como para humanistas em geral, a adoção de técnicas e algoritmos capazes de acelerar processos de identificação, classificação e organização de informações se tornou indispensável para o avanço no campo da história e historiografia das ciências.

Com alguns conceitos explorados sobre o modo como transpomos dados do meio físico para o meio digital, e como podemos representá-los imageticamente com uma narrativa não-linear, que relações podemos estabelecer entre a cognição humana e as capacidades de processamento lógico de computadores e algoritmos, notadamente no campo da aprendizagem de máquina e a inteligência artificial? São perguntas para abrir caminhos para essa proposição teórico-metodológica transdisciplinar, à medida que vemos cada vez mais o registro histórico, científico e de modo geral todo o conhecimento humano, sendo atravessado quase sempre por meios de comunicação digitais e instantâneos.

Abrangendo a multiplicidade de abordagens que temos adotado nessa pesquisa, a exploração transdisciplinar traduz um sentimento de política científica engajada para garantir que a descoberta de conhecimento seja livre a aberta à comunidade científica e aos cidadãos preocupados com a memória e a história das ciências no Brasil.

Um movimento de manifestação transformativa ou reveladora, que tem procurado transgredir as condições instauradas na pesquisa científica por uma herança que determinava a separabilidade dos saberes, dos humanos e o mundo natural. A transdisciplinaridade enquanto rompimento das fronteiras epistemológicas traz à luz a complexidade que ora se apresenta diante de problemas gerados pela rigidez disciplinar, revelando múltiplos níveis de realidade e a adoção de uma lógica não-clássica formando uma teia relacional complexa.

Enquanto um novo movimento de ideias, considerando também o próprio meio digital em agora que transita, a transdisciplinaridade invade a racionalidade

contemporânea como um manifesto de resgate do que é complexo, interligado e infinitamente relacional. (NICOLESCU, 2018). Será o digital um novo nível de realidade constituído de uma dimensão invisível, intuitiva, que traz em sua constituição os códigos necessários para traduzir percepções e comportamentos irracionais da nossa existência para com o mundo e a natureza? Talvez a investigação que se desenvolve aqui vá mais além da documentação histórica e científica. Talvez estejamos abrindo caminho para uma epistemologia mais integrativa, humanizada e livre.

3.2 O *ETHOS* DAS HUMANIDADES DIGITAIS: PENSAR, FAZER, SER

As Humanidades Digitais fizeram ressurgir questionamentos sobre um dualismo epistêmico do pensar e fazer, entre a teoria e o método. Uma questão de fronteiras científicas, característica recorrente nas discussões realizadas na História das Ciências.

Uma dessas fronteiras é a compreensão que construímos sobre o conceito de artefato, mais especificamente de artefatos intangíveis, levando-se em consideração o contexto e a caraterística como novidade na construção, preservação e partilha de conhecimento.

No caso do documento escrito, especialmente os documentos históricos e científicos, as humanidades digitais lançam um novo olhar sobre a mediação dos computadores como um fator de transformação; não apenas os efeitos da digitalização isoladamente, mas que transformações de significados e características esses documentos podem revelar ao serem transcritos para novos formatos.

Mídias tradicionais, como jornais, revistas, livros e periódicos, se transformam em algo novo pela mediação do computador. A partir da digitalização há uma quebra da essência linear de sua estrutura textual. Transgredindo a construção do autor original, dá-se uma nova função ao leitor: o de cartógrafo.

Roberto A. Busa (1913-2011) nos deu um quadro geral de suas reflexões (NYHAN & PASSAROTTI, 2019) acerca das perspectivas em humanidades computacionais já em meados do século XX. Para Busa, o campo era precisamente

a automação de toda análise possível das expressões humanas, no sentido mais amplo possível, da música ao teatro, do design e pintura à fonética, mas cujos núcleos se encontram no discurso de textos escritos.

Sua contribuição histórica no processamento de dados textuais congrega três diferentes perspectivas: a miniaturização tecnológica da informação (transcrição de informações de antigos formatos para mídias que armazenam informação em menor espaço, sendo este físico ou digital); a digitalização (como mencionado anteriormente), a informatização textual (levando a esforços no campo do armazenamento documental, especialmente banco de dados e a própria internet), a editoração de mídias eletrônicas (e suas potencialidades de transmissão de informação) e uma hermenêutica computacional como processo de organização e indexação de dados linguísticos completos (o que posteriormente deu origem às ferramentas de tradução automatizada).

Desde sua gênese, as humanidades computacionais, e mais recentemente as humanidades digitais, foram acompanhadas de avanços tecnológicos que permitiam a expansão de suas fronteiras epistemológicas e metodológicas com o foco de preservar a diversidade da atividade humana e nossa capacidade de registrar nossa história. O campo tornou possível a representação e a ressignificação de nossos registros históricos para que revelassem características e pistas não tão evidentes quando em seu estado textual original.

Isso levou a elaboração de infraestruturas de pesquisa adaptadas a ambientes eletrônicos como um impulso natural para um gerenciamento e uma engenharia sistematizada do conhecimento dentro do campo. Infraestruturas também no sentido cognitivo, como uma mudança de paradigma no modo de pensar a construção de conhecimento, reconhecendo interlocuções entre conceitos, processos e projetos que representam a práxis metodológica originária dentro da comunidade de humanistas digitais e, espera-se, na comunidade científica como um todo. Interlocuções que se dão, quase sempre, por meio de ferramentas computacionais.

Sendo essa uma das preocupações teóricas da presente pesquisa: que narrativas construímos na interlocução entre sujeitos e interfaces computacionais na representação e descoberta de conhecimento histórico e científico? Levando em

consideração que tais ferramentas computacionais têm seus algoritmos desenvolvidos a partir de valores específicos da cognição humana, torna central a questão das implicações para a construção do conhecimento por intermédio de uma cognição automatizada por códigos pré-concebidos. No entanto, a aplicação de buscas e recuperação de informação e processos críticos facilitados por algoritmos, originados de trabalhos das ciências humanas, tem sido demonstrada para além das expectativas no contexto da gênese do campo.

Pensar novos modos de representação e descoberta de conhecimento histórico e científico não é, e talvez nunca seja uma tarefa fácil. Requer uma reorientação de nossos processos mentais relacionais, em que os primeiros passos sejam experimentais desde a formulação das bases ontológicas e epistemológicas.

É a partir desse ethos epistêmico, um modo operacional teórico-metodológico das Humanidades Digitais, que exploramos aqui a dimensão teórica do referido campo a partir de seu Estado da Arte.

3.3 O DIGITAL: UM (NEM TÃO NOVO) TERRITÓRIO DAS HUMANIDADES

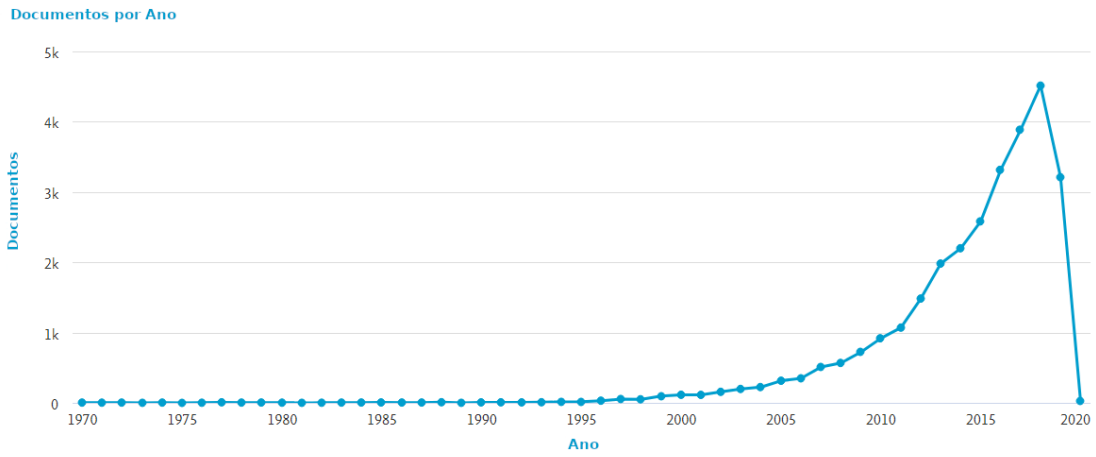
Um dos caminhos para compreender a gênese do campo das humanidades digitais é explorar o que tem sido escrito a respeito da formação de núcleos básicos de pesquisa, projetos e laboratórios experimentais, a elaboração de revistas e a publicação de obras dedicadas ao tema, assim como a elaboração de uma nomenclatura. Para elucidar um pouco mais a questão da formação do campo das Humanidades Digitais, tentaremos nos valer de recursos imagético-computacionais enquanto método de inquirição preliminar em conjunto com a base de dados Scopus.

Considerando que os termos “*humanities computing*”, “*humanist informatics*”, “*literary and linguistic computing*” e “*digital resources in the humanities*” podem ser compreendidos como termos que precederam o termo mais atual. A partir da pesquisa do termo “*digital humanities*”, desconsiderando qualquer filtro, obteve-se um resultado de 28.970 publicações indexadas até o ano de 2019.

Com um recurso de análise interna da própria base de dados, podemos verificar que as primeiras publicações datam do ano de 1970 (Figura 8). Passando por um

período de crescimento a partir dos anos 2000, atingindo por volta de 4500 publicações em 2018.

Figura 11: Ocorrências do termo “Digital Humanities” - Publicações/Ano - Indexação Scopus



Fonte: SCOPUS/Autor.

Em 1970 discutia-se sobre as potencialidades da utilização de recursos computacionais em disciplinas das humanidades. Em “*The Computer in the Humanities and Fine Arts”* (SEDELOW, 1970) investigou-se o uso de computadores aplicados nas artes e humanidades, indicando a relevância de pesquisas como o reconhecimento de padrões e análises em artes, arquitetura, música e literatura, como também a representação e manipulação de dados, significando uma potencial transformação do artefato físico em algo adequado à interpretação do computador.

Em 1971, já com uma certa atividade, o periódico “*Computers and the Humanities”* (1966-2004) publicava, sob autoria de George F. Putnam (PUTNAM, 1971) o impacto dos métodos quantitativos no trabalho de historiadores soviéticos e a direta influência do Partido Comunista Soviético na formulação de diretrizes epistemológicas para a disciplina histórica com estreita relação com métodos computacionais “para o desenvolvimento futuro das ciências sociais e o aumento de seu papel na construção do comunismo (…) e o avanço do conhecimento”. Ainda segundo Putnam, a questão do papel da História em consonância com a Computação foi explorada no artigo “*History, Concrete Social Investigations and Cybernetics”*

(POLYAKOV; POLETAEV; USTINOV *apud* PUTNAM, 1971) onde os autores apontavam que:

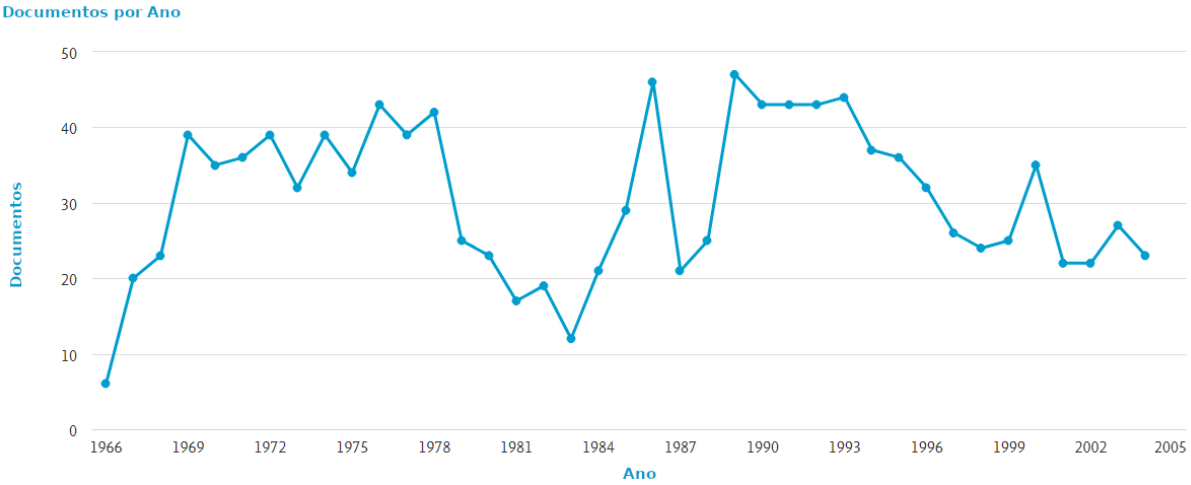
> ...ao usar ferramentas matemáticas e orientadas por computador, cientistas sociais estão avançando em direção a um objetivo científico fundamental, ao desenvolvimento de uma linguagem comum na base da qual frutíferos trabalhos interdisciplinares podem ser conduzidos e contribuições podem ser feitas em direção a solução de complexos problemas sociais na União Soviética.

Traçando caminhos relativamente diferentes de uma ciência ocidental, essa visão particular demonstra uma construção transdisciplinar dos saberes, agregando visões políticas ao empirismo científico. No contexto específico da Guerra Fria, aliar a construção de conhecimento à ação política poderia significar a liderança nos jogos de poder entre nações. Um esforço conjunto entre pesquisadores das humanidades, matemáticos e cientistas da computação no tratamento de fontes documentais em larga escala.

Outro historiador russo que agregou métodos estatísticos computacionais a seu trabalho foi Ivan D. Kovalchenko, cujo trabalho mais importante, "*The Russian Serf Peasantry in the First Half of the 19th Century*" (KOVALCHENKO, 1965), explorava as objeções à aplicabilidade da matemática às análises de dados históricos. Para Kovalchenko:

> ...historiadores não deveriam temer que o uso da linguagem universal da matemática trouxesse uma 'excessiva formalização da essência dos fenômenos históricos' (…) apenas aqueles elementos que podem ser quantificados são sujeitos ao tratamento matemático, e apenas como meios para explicar suas características qualitativas.

Figura 12: Publicações do Periódico “Computers and the Humanities” (1966-2004).



Fonte: SCOPUS/Autor.

Numa análise mais aprofundada do periódico “*Computers and the Humanities”* a partir das estatísticas da Base Scopus (Figura 9), podemos perceber grandes variações de produção de artigos. Totalizando 1.194, da sua criação até o encerramento de suas atividades, tendo o periódico dedicado suas publicações essencialmente à área das ciências sociais.

O artigo “*The Next Step”* (MILIC, 1966) já colocava em pauta as potencialidades dos métodos computacionais na interação com as disciplinas das humanidades, desde a análise de dicionários biográficos até bibliotecas automatizadas, mas reconhecendo, ainda assim, que o advento do computador mal tinha começado a ser completamente entendido. Mesmo porque, questionava-se uma dependência do acadêmico em relação ao programador na realização de projetos interdisciplinares no uso de computadores.

Para Milic,

> Nosso medo que o estudo da literatura possa se tornar mecânico, se for processado por computadores, tem impedido de tentarmos entender sua riqueza e possibilidade genuínas. A menos que tentemos entendê-lo como entendemos qualquer de nossas ferramentas, seremos não apenas incapazes de explorar seus recursos propriamente, como também estaremos em perigo de nos tornarmos suas vítimas. Controle vem com a compreensão, da fusão do usuário e o instrumento, como o braço e o sabre, como o cavaleiro e sua montaria. (MILIC, 1966, p. 4).

Um dos desafios colocados por Milic é o reconhecimento de padrões em narrativas: sejam padrões de escalas musicais, o uso de cores numa obra de arte, estruturas poéticas, ou mesmo processos epistêmicos relacionais a partir de artigos científicos. Numa escala pequena, a mente humana não tem problemas nesse reconhecimento de padrões, mas quando se trata de relativizar e comparar um corpo documental mais extenso e suas complexas dimensões referenciais, o uso de computadores por humanistas ainda hoje é um desafio. Não só um desafio cognitivo e analítico, como também formativo do campo.

Ainda segundo Milic,

> Ao contrário do observador humano, no entanto, o computador pode ser feito para detectar o padrão mais longo e mais oculto, independentemente da aparência aleatória que ele apresenta ao olho humano. Assim, devemos aprender a fazer perguntas maiores do que podemos responder e a detectar o que escapa aos nossos sentidos desassistidos. Isso pode envolver não apenas propor questões antigas de novas maneiras, mas também pensar em novas questões. O computador pode ser feito uma extensão do homem apenas se abrir caminhos pelos quais não suspeitamos a existência. (*op.cit*, p. 6).

Problema semelhante é exposto no artigo "*The Computer and the Historian"* (BULLOUGH, 1967), em que se verifica uma resistência por parte da comunidade de historiadores sobre as implicações no uso de computadores, como a possibilidade de estabelecer a importância de fatos passados de modo preditivo. Mas reconhece-se que, à medida que novas técnicas e métodos se tornam parte da disciplina, há uma necessidade por conhecimento e treinamento mais especializado. Considerando que o trabalho do historiador é visto como altamente individualista, onde o sujeito está imerso em seu objeto, a adoção de práticas exógenas requer colaboração transdisciplinar cada vez maior e mais significativa.

Também a estilometria e os estudos de atribuição ganharam poder de processamento e análise com a mediação dos computadores. Em "*The Computer and Plato's Seventh Letter"* (MORTON, 1967) explora-se as possibilidades de identificação de autoria a partir da extração de palavras chave, repetição no uso específico de

expressões e também a extensão de sentenças, levando a criação de um banco de dados de estilo para estudos comparativos.

Incursões analíticas nos “Diálogos de Platão” formaram os princípios da estilometria, tendo como seu precursor o filósofo polonês Wincenty Lutoslawski, seguido mais tarde por Frederick Mosteller e David Wallace nas análises de autoria nos Arquivos Confederados, um conjunto de 85 artigos escritos por Alexander Hamilton, James Madison e John Jay sob o pseudônimo de “Publius”. Para além das questões de autoria, a estilometria com mediação de computadores pode revelar novos insights no campo da ciência da informação, como também em estudos em documentos históricos com técnicas forenses associadas à inteligência artificial.

No artigo “*The Thomas Mann Project”* (LEHNERT; RUECKING; PORTER, 1967) expõe-se o problema da transcrição de textos físicos para mídias que possam ser lidas por computadores. Foi necessário a criação de um índice para que o processamento por computador pudesse gerar uma resposta automatizada e mais eficiente. Isso demonstra que, ainda que nos utilizemos dos computadores como extensores cognitivos, é necessário um esforço organizacional e de construção linguística por parte de humanos para que esse processo seja realizado pela máquina; uma simbiose condicional.

Para historiadores tradicionais, o uso de computadores para processamento de dados históricos ainda é um grande desafio, haja vista a capacidade de memória e análises estatísticas que a máquina é capaz de realizar de modo diferente do sujeito humano. No entanto, os avanços nas tecnologias têm dado maiores garantias na formulação de conclusões baseadas em análises sistemáticas associadas a dados históricos relevantes.

Garantias que servem para quantificar a veracidade das evidências históricas e, a partir disso, efetivamente, demonstrar que a computação tem representado um progresso teórico-metodológico considerável para a disciplina histórica. Mas não podemos esquecer que, inicialmente, os fenômenos, fatos e dados históricos fazem parte de um campo cuja subjetividade humana atravessa seus objetos de pesquisa, o que torna a mensuração ainda um grande dilema.

Ao elaborar questões para a ciência histórica estamos estabelecendo certas fronteiras na busca pela identificação, classificação e significação de nossos objetos, que podem ser medidos a partir de valores qualitativos ou quantitativos. No entanto, não importa apenas quantas questões levantemos, mas também o que esses processos cognitivos, ontológicos e epistemológicos de elaboração significam. E ao usar o computador, esses processos transdisciplinares, por vezes indisciplinados, podem descortinar questões que a ciência histórica, num contexto metodologicamente tradicional, não faria.

Desse modo, a disciplina histórica pode se valer de propostas teórico-metodológicas a partir de questões "orientadas à problemas", para usar um termo comum à computação, e avançar no trato de dados, informações e fenômenos históricos pela mediação do computador na resolução de antigas ou mesmo novas perguntas. A exemplo do porquê, como e o que fazer com a imensa quantidade de dados gerados a partir de um reprocessamento de coleções de documentos históricos.

Diz-se aqui reprocessamento pois já avançamos nas metodologias de arquivamento digital em que a infraestrutura tecnológica da Internet exerce um grande papel na preservação e disseminação de documentos históricos. É preciso ir mais além e saber o que podemos aproveitar dessa infraestrutura e das potencialidades mediadas por computador. Afinal, o elemento mais importante dessa proposta transdisciplinar é o método, visto que o computador tem maior potencial de processamento para avaliar a importância relativa das diferentes forças no contexto de eventos históricos documentados, ou seja, exerce a função precisa e sistemática de medir uma grande quantidade de dados (SILBEY, 1972).

Em 1975, no artigo "*Building a Data File from Historical Archives"* (LABARRE; LEVITT, 1975) discutia-se a construção de banco de dados históricos a partir da extração de informações de documentos. Nesse contexto, como ainda hoje, dois grandes problemas são encontrados no processamento de dados históricos documentais: transformar os dados para um modelo compatível para análise computacional, e o desenvolvimento de técnicas para análise desses dados. Com exceção bastante restrita das séries históricas de dados econômicos, historiadores

não tem familiaridade com dados e métodos estatísticos, haja vista a incompatibilidade com a subjetividade inerente aos fenômenos históricos.

Assim, a constituição de bancos de dados históricos computáveis, devidamente quantificados e classificados, devem prover não apenas informações quantitativas relevantes para a análise de relações dos fatos e narrativas históricas, mas também preservar o caráter de cada elemento informacional histórico para que sejam retomados seletivamente e outras relações, de outro modo imperceptíveis, possam ser descobertas.

Ainda segundo Labarre e Levitt, historiadores dependeriam de uma forma de entrada de dados compatíveis com os dados históricos em seu formato original, fossem manuscritos ou com letra de máquina. O desafio para historiadores é, ainda hoje, a inserção, quase sempre manual, e a manipulação de determinados dados históricos e suas análises posteriores com a mediação de computadores.

Uma das funções dos historiadores tem sido sempre a descoberta de relações individuais entre pessoas, objetos e eventos, e mais recentemente tem sido a de descobrir relações estatísticas na sociedade como um todo. Um arquivo computável de dados históricos, que preserve elementos descritivos desses dados assim como a informação quantitativa relacionada a eles pode melhor auxiliar historiadores nos seus trabalhos de pesquisa. Uma forma de inserção de dados que seja natural aos historiadores e que lhes permita capturar aqueles aspectos relacionais essenciais dos dados e, portanto, minimizando os erros inerentes a um processo de codificação numérica manual, é igualmente importante. (LABARRE & LEVITT, 1975. p. 82).

A transcrição de documentos históricos de seu formato original para o formato digital ainda é passível de erros, tanto na grafia como na verificação em paralelo com outros documentos que tratam dos mesmos dados, a exemplo de tabelas com dados econômicos elaboradas a partir de registros manuais de transações comerciais. Documentos históricos com grafias antigas podem representar um esforço maior na construção de uma base de dados que tem o reconhecimento linguístico como premissa para a identificação de relações entre entidades textuais.

A exemplo da língua portuguesa escrita durante o século XIX, palavras como “pharmacia” e “farmácia” devem ser equiparadas para que a análise textual seja

eficiente e ampla. Metodologias de transcrição permitem a inserção de dados em aplicações de computador especializadas na construção de grafos e redes relacionais temáticas de acordo com o que é tratado no documento ou num grande conjunto de documentos institucionais, por exemplo.

Levando, assim, a uma cartografia linguística, computacional e imagética dos documentos e dos conhecimentos históricos e científicos. Cartografia que deve se valer também de um banco de dados semântico, que leve em consideração conjuntos de expressões que podem revelar associações a outros elementos, em outros documentos, que complementem as informações constadas naqueles.

Um valor, uma ideia ou padrão de comportamento tem seu próprio movimento? Ou há uma força que deriva apenas de instituições formalizadas? (...) Esse é outro sentido de uma era pós-histórica: não que o passado não nos alcance, mas aqueles elementos estruturais subjacentes evaporam num turbilhão de mudanças que nenhuma base identificável permanece para nossa vida coletiva. (...) Primeiro, para poder responder a questões sobre processos históricos, temos primeiro que ter informações precisas e adequadas sobre a história de nossa civilização ao longo de um tempo significativo; segundo, precisamos de habilidades para processar essas informações, analisa-las e revelar interconexões, estruturas ocultas e características temporais inerentes; terceiro, temos de integrar os resultados de nossas análises de diferentes áreas para responder questões complexas. (THIBODEAU, 1976. p. 89).

No breve artigo "*Machine Readable Archives and Future History"* (1976) Kenneth Thibodeau lança questionamentos sobre os esforços necessários de Historiadores, no futuro, para identificarem os processos de mudança social, se profundas ou superficiais. Primeiramente, a aquisição de massas de dados apropriadamente organizados e coerentes; segundo, a formalização de medições objetivas e inequívocas; terceiro, a habilidade de extrair dos dados as informações necessárias para aplicar as medidas de mudança e executar os testes associados a estas.

Para determinar as relações lógico-semânticas, uma enciclopédia ou rede semântica é frequentemente sugerida para identificar o escopo e extensão de um

dado campo do conhecimento, as relações conhecidas entre conceitos incluídos no campo.

As relações linguístico-semânticas, por sua vez, são obtidas usando uma análise sintática-semântica combinada para gerar uma estrutura detalhada dos textos escritos. Quando o conhecimento mundial (enciclopédico) é combinado com a análise linguística, um texto deve ser representável como uma série de “cadeias de inferência” que representam a linha de pensamento expressa no texto, incluindo as condições e premissas declaradas e não declaradas”. (SALTON & WONG, 1976. p. 69)

Nesta triangulação das problemáticas envolvendo o uso da computação na análise de documentos históricos cabe ressaltar como essa massa de dados são (e serão) produzidas dentro de um contexto metodológico formativo de historiadores profissionais, incluindo uma discussão sobre currículo transdisciplinar para a História das Ciências, uma temática a ser explorada mais adiante.

Com a mediação por computadores, nosso entendimento sobre o conceito de documento também se ampliou. Sendo considerado um meio físico, o documento textual impresso possui características físicas peculiares ao seu meio material, seja o papel, o papiro, tabuletas de argila ou qualquer outro tipo de meio registrado na nossa história. Agora, em meio digital, o documento textual passa a apresentar novas possibilidades estruturais na sua leitura e em sua manipulação, facilitadas pelo processo de binarização e digitalização.

No artigo “*Multi-Dimensional Texts in a One-Dimensional Medium”* (HUITFELDT, 1994) são exploradas características topográficas (colocação que será melhor explicada ao final deste capítulo) de textos transcritos para o meio digital pela mediação do computador. Trata-se de uma análise proposicional sobre as potencialidades do texto digital e as similaridades linguísticas, ontológicas e estruturais com documentos textuais em mídias tradicionais. Para Huitfeldt, a adaptação de conceitos tradicionais sobre texto para uma nova mídia pode ajudar a explicar tal processo de construção conceitual desse objeto linguístico, como também seu processo de codificação do meio físico para o digital.

Huitfeldt posiciona-se de modo a considerar que textos são entidades linguísticas multidimensionais, em que a capitalização e as pontuações são elementos

de fronteira, enfileirados de modo coeso para a construção de uma narrativa. É posto uma questão de representação da informação textual nos dois meios, físico e digital: como mapear e transcrever toda a informação para o computador, para o meio digital? Daí surge a necessidade de elaboração, formalização e padronização de códigos textuais específicos para o novo meio.

O propósito da codificação textual descritiva não é preparar para algum modo específico de apresentação ou análise, mas para representar o mais fiel possível a informação textual, a estrutura lógica do texto e as relações internas entre diferentes segmentos de texto.” (HUITFELDT, 1994. p. 236).

Cabe aqui uma discordância pontual com a visão de Huitfeldt. A transcrição de textos do meio físico para o meio digital se dá com o propósito de análise e (re)presentação, visto que as potencialidades linguísticas do meio digital fazem com que o texto transcrito tenha codificado em si metadados relacionais acerca do contexto que o constitui, isso inclui dados sobre autor, data, número de páginas. Ademais, a partir de processos de extração de informações textuais do documento digital é possível construir uma cartografia de entidades específicas, como a identificação de nomes próprios, lugares, referências e citação cruzada. A bibliometria ganha uma nova dimensão digital, que é a webometria.

Assim, tornava-se premente a necessidade de exploração das ideias e técnicas correntes do campo da ciência da computação para elevar princípios e modelos de processamento documental ao mesmo nível das linguagens de programação e construção de banco de dados.

Ao apontar uma multidimensionalidade textual comum aos dois meios Huitfeldt põe em xeque a hierarquização linear presente apenas nos documentos textuais físicos, enquanto que no novo meio digital, as possibilidades de representação da informação vão além de segmentos textuais enfileirados e delimitados por pontuações.

Aqui, a proposta de uma cartografia teórico-metodológica, tendo como fonte um documento histórico e científico, possibilitada pela mediação do computador, visa justamente confirmar o que Huitfeldt apontou como uma superação de uma dimensão única, linear e visível do documento textual.

A representação computacional-imagética da informação só tem sido possível devido aos avanços nas potencialidades de codificação e manipulação da informação em meio digital. Sem desconsiderar a importância pela preservação dos documentos originais em estado físico, não só por uma política científica de valorização da memória, mas também como um banco de dados referencial pela verificação e garantia das verdades históricas contidas nesses documentos. É o surgimento de uma nova geração de humanistas digitais, de teorias e métodos pela preservação documental da memória histórica e científica.

O cenário brasileiro de desenvolvimento teórico, metodológico e experimental no campo das humanidades digitais também é bastante expressivo. Dentre estes, o MediaLab[19], da Escola de Comunicação da Universidade Federal do Rio de Janeiro, tem suas pesquisas "voltadas para a intersecção entre tecnopolíticas, subjetividades e visibilidades". Explorando métodos digitais de análise e visualização de dados no campo das humanidades, experimentando diferentes linguagens, metodologias e perspectivas conceituais na produção e divulgação de pesquisas, permeando as urgências (e emergências) políticas e sociais do presente.

O Laboratório de estudos sobre Imagem e Cibercultura[20] (Labic), do Departamento de Comunicação Social da Universidade Federal do Espírito Santo, é associado ao Programa de Pós-Graduação em Comunicação e Territorialidades e também ao Programa de Pós-Graduação em Ciência da Computação. O laboratório tem como missão a realização experimental de produtos digitais e a promoção de pesquisas e atividades de extensão relacionados ao impacto da cultura digital nos processos e práticas de comunicação contemporânea.

Enquanto comunidade de práticas, por fim, temos o Laboratório em Rede de Humanidades Digitais[21] (LARHUD) como parte de um programa de "transformação digital" da pesquisa aplicada em humanidades, desenvolvendo ferramentas e metodologias mediadas por plataformas digitais, usando o conceito de *thinkering*

---

[19] MediaLab UFRJ. Ver mais em: https://medialabufrj.net/sobre/

[20] Labic UFES. Ver mais em: https://www.labic.net/labic

[21] LARHUD IBICT. Ver mais em: http://www.larhud.ibict.br

agregando as atividades de investigação, pesquisa, inovação e ludicidade enquanto um espaço de práticas científicas.

3.4 O DOCUMENTO COMO TERRITÓRIO: VAMOS CARTOGRAFAR!

Como processo final do estabelecimento da dimensão teórica da pesquisa, a proposição de um mapeamento documental vem oferecer uma nova visão espacial e imagética sobre a construção de conhecimento histórico e científico e como tal dimensão pode revelar novas narrativas diante das possibilidades e potencialidades do campo transdisciplinar das Humanidades Digitais.

Considerando que cada documento histórico e científico tem um potencial gerador de narrativas, em que cada um é também continuidade, rompimento, estrapolação e convergência de ideias, o modo de tratamento teórico-metodológico adotado para a exploração do conceito de cartografia advém do conceito deleuziano adotado por KASTRUP, PASSOS e ESCÓSSIA (2010) na análise de processos de produção de subjetividades. Ao que chamam de Pistas do Método da Cartografia, são identificados processos de formação de subjetividades com características singulares na construção de um determinado campo de atuação.

Segue-se uma descrição sumária de seis pistas sobre o método da cartografia:

Cartografia como Pesquisa-Intervenção

Não há uma prescrição metodológica ou objetivos pré-estabelecidos. Reversão do sentido tradicional de “método”, não para alcançar uma meta, mas para descobri-las ao percorrer o território. Considera os efeitos do processo de pesquisar sobre o objeto de pesquisa, sobre o pesquisador e seus resultados. O ponto crucial dessa pista é a experiência do “saber-fazendo”. É inverter a direção da pesquisa, do conhecer-transformar a realidade, para o transformar-conhecer a realidade. “É a indissociabilidade entre o conhecimento e a transformação, tanto da realidade quanto do pesquisador.” (KASTRUP *et al.*, V. 2010, p. 14).

A Atenção do Cartógrafo

A cartografia como um processo é sempre um método *ad hoc*. Diálogo entre a atenção flutuante (Freud) e o reconhecimento atento (Bergson). Analise das etapas iniciais da pesquisa. De um ponto de vista construtivista, não há coleta de dados, mas sim uma produção de dados. Quatro são os gestos da atenção cartográfica: o rastreio, o toque, o pouso e o reconhecimento atento. A atenção do cartógrafo não é a de simples seleção de informações. Sua atenção pode assumir diferentes funcionamentos, dependendo das políticas cognitivas em ação. “A pergunta, que diz respeito ao momento que precede a seleção, seria melhor formulada se evidenciasse o problema da própria configuração do território de observação” (KASTRUP *et al*., 2010. p. 35)

Cartografar é Acompanhar Processos

Cartografar é um modo de fazer, anuncia uma prática. A criação de elos. “A pesquisa cartográfica consiste no acompanhamento de processos, e não na representação de objetos.” (KASTRUP. *et al*, 2010. p. 53). Crítica à Ciência Moderna pelo fato de que a diferença entre sujeito e objeto não ser filosófica e sim prática, não é uma categoria transcendental, mas uma configuração histórica. O método da cartografia é uma resposta à recusa pelo reconhecimento do caráter inventivo das práticas científicas em produzirem conhecimento. A pesquisa de campo requer a habitação de um território que, em princípio, ele não habita. É necessária uma observação participante, um agenciamento.

A Cartografia Processual não visa isolar o objeto de suas articulações históricas nem de suas conexões com o mundo. Seu objetivo é desenhar a rede de forças à qual o objeto ou fenômeno em questão se encontra conectado, dando conta de suas modulações e de seu movimento permanente. (KASTRUP *et al*, 2010, p. 57)

O Coletivo de Forças como Plano da Experiência Cartográfica

"...Ao lado dos contornos estáveis do que denominamos formas, objetos ou sujeitos, coexiste o plano coletivo das orças que os produzem, além de definirem a cartografia como prática de construção desse plano." (KASTRUP *et al*, 2010, p. 15).

A cartografia visa a ampliação de nossa concepção de mundo para incluir o plano movente da realidade das coisas. O conceito de coletivo proposto pela cartografia processual é entendido como a relação entre o plano das formas e o plano das forças; planos não distintos, assegurando relações de reciprocidade e cruzamentos múltiplos. Os objetos do mundo, segundo essa visão processual, se distanciam de uma natureza fixa e invariante, ao contrário, se abrem às variações e estão em constante processo de transformação.

"Ao optar pelo uso do termo plano de forças e não campo de forças, buscamos o afastamento da tradição iniciada no século XIX e que inspira as ciências naturais e as ciências humanas a pensar os fenômenos como resultantes do equilíbrio da dinâmica de forças." (KASTRUP *et al*, 2010, p. 96).

Cartografar é Habitar um Território Existencial

"O ato cognitivo, base experiencial de toda atividade de investigação, não pode ser considerado, nesta perspectiva, como desencarnado ou como exercício de abstração sobre dada realidade." (KASTRUP *et al*, 2010, p. 131). O trabalho da cartografia não pode se fazer como sobrevoo conceitual sobre a realidade investigada, é sempre pelo compartilhamento de um território existencial que sujeito e objeto da pesquisa se relacionam e se codeterminam. É uma expressividade, e não uma funcionalidade, que explica a formação territorial (DELEUZE & GUATTARI, 1995). O território é uma assinatura expressiva que faz emergir ritmos como qualidades próprias que, não sendo indicações de uma identidade, garantem a formação de certo domínio. (KASTRUP *et al*, 2010, p. 133).

> O território não se constitui como um domínio de ações e funções, mas sim como um ethos, que é ao mesmo tempo morada e estilo. Os sujeitos, os objetos e seus comportamentos deixam de ser o foco da

> pesquisa, cedendo lugar aos '"personagens rítmicos" e às "paisagens melódicas". (KASTRUP *et al*, 2010, p. 134).

A instalação da pesquisa cartográfica sempre pressupõe a habitação de um território, o que exige um processo de aprendizado do próprio cartógrafo.

> Segundo a perspectiva cartográfica, a construção de um território existencial não nos coloca de modo hierárquico diante do objeto, como obstáculo a ser enfrentado (conhecer = dominar, objeto = o que objeta, o que obstaculiza). Não se trata, portanto, de uma pesquisa sobre algo, mas uma pesquisa com alguém ou algo. Cartografar é sempre compor com o território existencial, engajando-se nele. (KASTRUP *et al*, 2010. p. 135).

Por uma Política da Narratividade

A alteração metodológica proposta pela cartografia exige uma mudança nas práticas de narrar. "Os dados coletados a partir de diferentes técnicas (...) indicam maneiras de narrar, seja dos participantes ou sujeitos da pesquisa, seja do pesquisador ele mesmo, que apresentam os dados, sua análise e suas conclusões segundo certa posição narrativa." (KASTRUP *et al*, 2010, p. 150).

A escolha da posição narrativa não pode ser encarada como desarticulada das políticas que estão em jogo: políticas de pesquisa, políticas da subjetividade, políticas cognitivas. Toda produção de conhecimento se dá a partir de uma tomada de posição que nos implica politicamente.

> Traçar o transversal é, no que diz respeito aos modos de dizer, tomar a palavra em sua força de criação de outros sentidos, é afirmar o protagonismo de quem fala e a função performativa e autopoiética das práticas narrativas. Um poder coletivo em detrimento das inibições individuais, uma atenuação do medo de ser massacrado, asfixiado, em razão de uma transgressão que se passa no nível das cadeias significantes inconscientes. (...) Transgredir, nesse sentido, é enfrentar as práticas de assujeitamento/subjetivação assentadas no medo imposto por um significante social que opera psiquicamente como instância judicativa ou superegóica" GUATTARI *apud* KASTRUP *et al*, 2010, p. 156-157).

E segue,

> Transgredir não pode mesmo ser pensado sem a sua inevitável relação com o que, do ponto de vista psicopatológico, se define como perversão. O perverso frente ao imperativo da castração transgride o

> medo, locupletando todo e qualquer sinal de falta enquanto imperativo legal. (...) A transgressão como método se faz imediatamente também como transgressão do método (...) nos afasta das exigências de purificação, de rigor asséptico que distingue e separa o fazer e o dizer. (KASTRUP *et al*, 2010, p. 156-157).

No método da cartografia, a inseparabilidade entre pesquisa e intervenção desestabiliza pressupostos tradicionais do conhecimento científico e o ideal de inteligibilidade que se hegemonizou como positivo, rigoroso, neutro, objetivo. (KASTRUP *et al*, 2010, p. 172).

Ao fim da breve descrição dessas pistas sobre o método da cartografia processual cabe justificar o uso análogo frente às demandas da pesquisa tanto na construção da dimensão teórica como da dimensão metodológica e experimental. Aplicar uma transversalidade de conceitos da psicologia clínica na construção de uma nova narrativa entre humanos e máquinas acaba por fazer sentido em novos modos de construção do conhecimento histórico e científico, tendo como base um conjunto de documentos em diálogo com diretrizes algorítmicas computacionais, ambos objetos da mediação humana.

Assim, é chegado o momento de executar políticas de transgressão para a construção de novos modos de fazer e saber ciências a partir de perspectivas transdisciplinares das Humanidades Digitais.

4 DIMENSÃO EXPERIMENTAL: REPOSITÓRIOS CIENTÍFICOS E SUAS POTENCIALIDADES METODOLÓGICAS

Inegavelmente a função de registrar, analisar, interpretar e mesmo escrever a história tem sido pesarosa. Desde Heródoto, passando por Marc Bloch, até os dias de hoje, especificamente hoje, às 17:58 do dia 18 de março de 2021, historiadores tem a função laboral da análise documental em que pese a classificação, organização e outros inúmeros processos de seleção de uma literatura factual na construção de uma narrativa que se entenda como importante em cada contexto de pesquisa.

E é no contexto espaço-temporal do Brasil no séc. XXI que se inicia aqui um estudo de caso para proposição metodológica enquanto política científica para a análise de documentação histórica, especificamente o periódico *Archivos do Museu Nacional*, com o objetivo de explorar novos modos de fazer e escrever ciências, novos modos de representar e descobrir conhecimento. Mas não sem antes nos confrontar com essas potencialidades da tecnologia que vem condicionando a função de historiadores há mais de meio século, de modo nem sempre uniforme, mas que avança na direção de uma transdisciplinarização dos métodos da História enquanto disciplina.

4.1 O TERRITÓRIO DO HISTORIADOR FRENTE OS DOMÍNIOS DAS HUMANIDADES DIGITAIS

Nessa perspectiva de encarar a documentação textual como um espaço de investigação, e que para tanto, é necessário um conjunto de métodos, técnicas e ferramentas, a ascensão da computação como um meio utilitário é abordado em “*The Territory of the Historian”* (LADURIE, 2017. p. 3) reforçando uma transformação do que se chamou de a “Revolução Quantitativa na História” ao estabelecer um novo campo, a Demografia Histórica ou História Demográfica, que por sua vez tem se desenvolvido para uma História Digital, um campo derivado das Humanidades Digitais, preocupado particularmente com análises de registros públicos mediados por tecnologias computacionais, registros esses que incluem fontes primárias escaneadas

e/ou digitalizadas, visualizações de dados históricos, mapas interativos, linhas de tempo, arquivos de áudio e mesmo a criação de espaços virtuais em 3D para representação histórica.

Não é preciso repetir aqui o que já entendemos sobre a complexa função do computador, como um aparato tecnológico que nos auxilia nas mais diversas tarefas da vida cotidiana, e mesmo seu uso no contexto da atividade acadêmica e científica, no entanto, como tem sido reforçado ao longo desta pesquisa, tem sido imprescindível compreender como essas novas linguagens, novas técnicas de análise documental e lógica interpretativa estão se entrelaçando também na prática historiográfica.

Cada vez mais é demandado de historiadores que saibamos também programar, nem sempre no sentido estrito, mas com a adoção de *softwares* especializados para a condução não só da pesquisa, como também da produção dos dados pesquisados. Pode parecer estranho falar em produção dos dados quando nos referimos às fontes históricas, normalmente estas já dadas como estáticas, cabendo a nós apenas interpretá-las. De certo modo isso está correto, no entanto, com a mediação de novas técnicas e tecnologias computacionais, tem sido possível criar fontes primárias, com uma natureza digital, flexível, manipulável.

Uma iniciativa recente chamada "*The Programming Historian*"[22] visa oferecer meios para que humanistas possam se valer de técnicas, linguagens de programação e *softwares* especializados na realização de tarefas hoje vistas como cotidianas para a pesquisa documental. Desde a realização de OCR em textos históricos, como a aplicação de algoritmos de classificação textual, são exemplos de metodologias oferecidas pelo grupo de pesquisadores responsáveis pela plataforma, que já conta com versões em espanhol, francês e em português.

A possibilidade de inovar as metodologias e acesso dos acervos e coleções de objetos que, apesar de todas as técnicas e ferramentas destinadas à sua análise e preservação, não deixaram de habitar um lugar escuro e inacessível no imaginário do público, e mesmo dos acadêmicos, tem sido possível com o armazenamento em nuvens e com o auxílio de inteligência artificial para o tratamento computacional do

22 https://programminghistorian.org/

material em grande escala. Vale lembrar aqui uma referência às tecnologias usadas no audiovisual, especialmente no cinema, em que muitos filmes só foram possíveis de serem produzidos quando surge o potencial tecnológico para tal tarefa[23].

Com os acervos históricos, não tem sido diferente: apenas depois de desenvolvidas determinadas tecnologias é que estamos podendo retirar os acervos da escuridão e os colocando sob a ótica das tecnologias computacionais a fim de oferecer novos caminhos para a descoberta de conhecimento, e é disso que se trata esse capítulo da tese.

4.2 O PRIMEIRO CONTATO: NAVEGANDO PELOS *ARCHIVOS DO MUSEU NACIONAL*

Ao acessar o site do Museu Nacional pela primeira vez em busca do acervo digital, por volta de 2015[24], foi possível verificar o acervo quase completo do periódico *Archivos do Museu Nacional*[25], resultado do gradual processo de implementação do projeto de preservação documental idealizado pela SEMEAR/MN/UFRJ, hoje gerido pela Comissão de Publicações.

Tomo como exemplo para o estudo de caso o primeiro volume do periódico, publicado em 1876, disponível integralmente em formato PDF escaneado. Nesse primeiro contato, se pode verificar um dos maiores desafios que historiadores enfrentam com documentação textual: a procura por termos relevantes se torna tão custosa como se tivéssemos de posse do documento físico, ou seja, mesmo com um processo inicial do que costumamos chamar de digitalização, na verdade nos deparamos apenas com uma versão fotocopiada por meio de processos de captura

23 Avanços Tecnológicos no Cinema. TV Gazeta. Disponível em: <https://www.tvgazeta.com.br/videos/avancos-tecnologicos-no-cinema-fevereiro-2020/>. Acesso em: 22 de fevereiro de 2021.

24 Tive o privilégio de ser estagiário bolsista no Museu Nacional atuando no Projeto “Popularizando a História do Brasil no Museu Nacional”, coordenado por minha orientadora acadêmica no HCTE/UFRJ, profa. Regina Dantas, atuante na SEMEAR/MN/UFRJ.

25 Sobre o assunto, ver: Paulo Vinicius Aprígio da Silva, O Archivos do Museu Nacional e a Institucionalização das Ciências no Século XIX, 2012.

de imagem, em que as informações contidas no texto original não são indexadas de modo que haja reconhecimento dos caracteres representados.

Assim como foi abordado no capítulo sobre a dimensão metodológica, coloco em prática uma cartografia processual na análise do documento, buscando um novo modo de representação do conhecimento contido na fonte primária e quiçá novas narrativas, à medida que as informações são extraídas e ressignificadas por meio de processos computacionais-imagéticos.

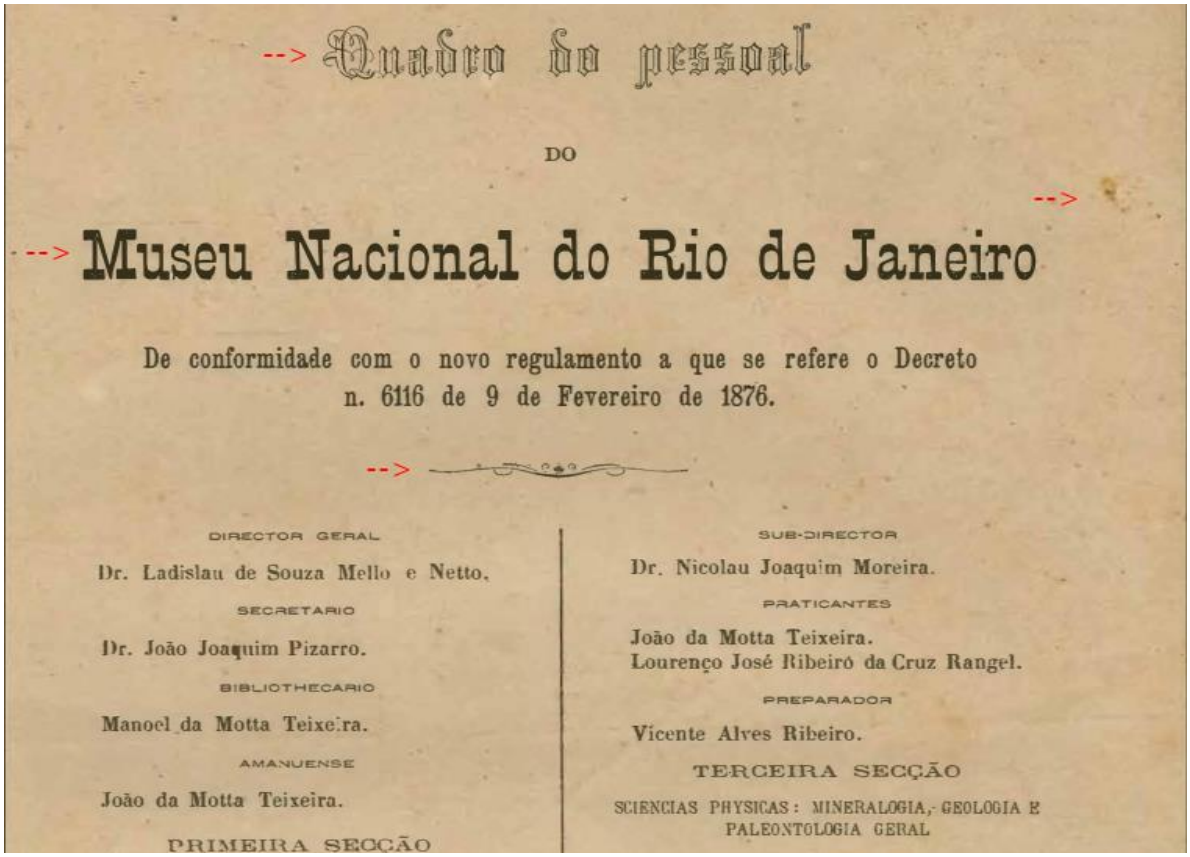
Assim, retomo o que considero a etapa essencial na análise de documentos históricos que passam por um processo digital, qual seja, a aquisição das imagens por meio de escaneamento, levando em consideração que a extração de dados se dá efetivamente apenas com a aplicação do reconhecimento óptico de caracteres (OCR).

Para tanto, o documento ainda precisa de etapas preparatórias, como por exemplo a verificação da qualidade da imagem escaneada, visto que imagens em alta resolução tem um resultado mais eficiente em processamento posterior. Devendo ser verificado também o alinhamento dos caracteres na imagem capturada. A depender dos equipamentos utilizados, a captura ocorre em condições adversas, com baixa luminosidade e de forma manual, onde as páginas não podem ser separadas do documento como um todo, o que ocasiona um abaulamento da superfície, dificultando uma imagem plana como resultado.

No caso da versão digital do primeiro volume dos *Archivos*, a captura foi realizada de maneira bastante satisfatória, permanecendo com a tonalidade original e bom alinhamento, no entanto, não foi identificado um processo de OCR e verificação ortográfica completa que permitisse uma análise textual exata quando o documento é colocado sob análise em *softwares* específicos para essa tarefa.

Abaixo (Figura 10) temos um recorte do documento escaneado em seu estado original, permitindo uma leitura linear do documento. Ainda assim, deve-se pontuar que certos elementos presentes na imagem, indicados em vermelho, podem oferecer obstáculo no que diz respeito aos primeiros processos de digitalização do documento. Mais adiante, iniciamos propriamente a inserção do documento em *software* especializado no tratamento de documento textual para aquisição de informações.

Figura 13: Recorte de documento destacando elementos que apresentam dificuldades no processamento do OCR, por se tratarem de manchas, gráficos ou tipografias incomuns.



Quadro do pessoal

DO

Museu Nacional do Rio de Janeiro

De conformidade com o novo regulamento a que se refere o Decreto n. 6116 de 9 de Fevereiro de 1876.

DIRECTOR GERAL

Dr. Ladislau de Souza Mello e Netto.

SECRETARIO

Dr. João Joaquim Pizarro.

BIBLIOTHECARIO

Manoel da Motta Teixeira.

AMANUENSE

João da Motta Teixeira.

PRIMEIRA SECÇÃO

SUB-DIRECTOR

Dr. Nicolau Joaquim Moreira.

PRATICANTES

João da Motta Teixeira.
Lourenço José Ribeiro da Cruz Rangel.

PREPARADOR

Vicente Alves Ribeiro.

TERCEIRA SECÇÃO

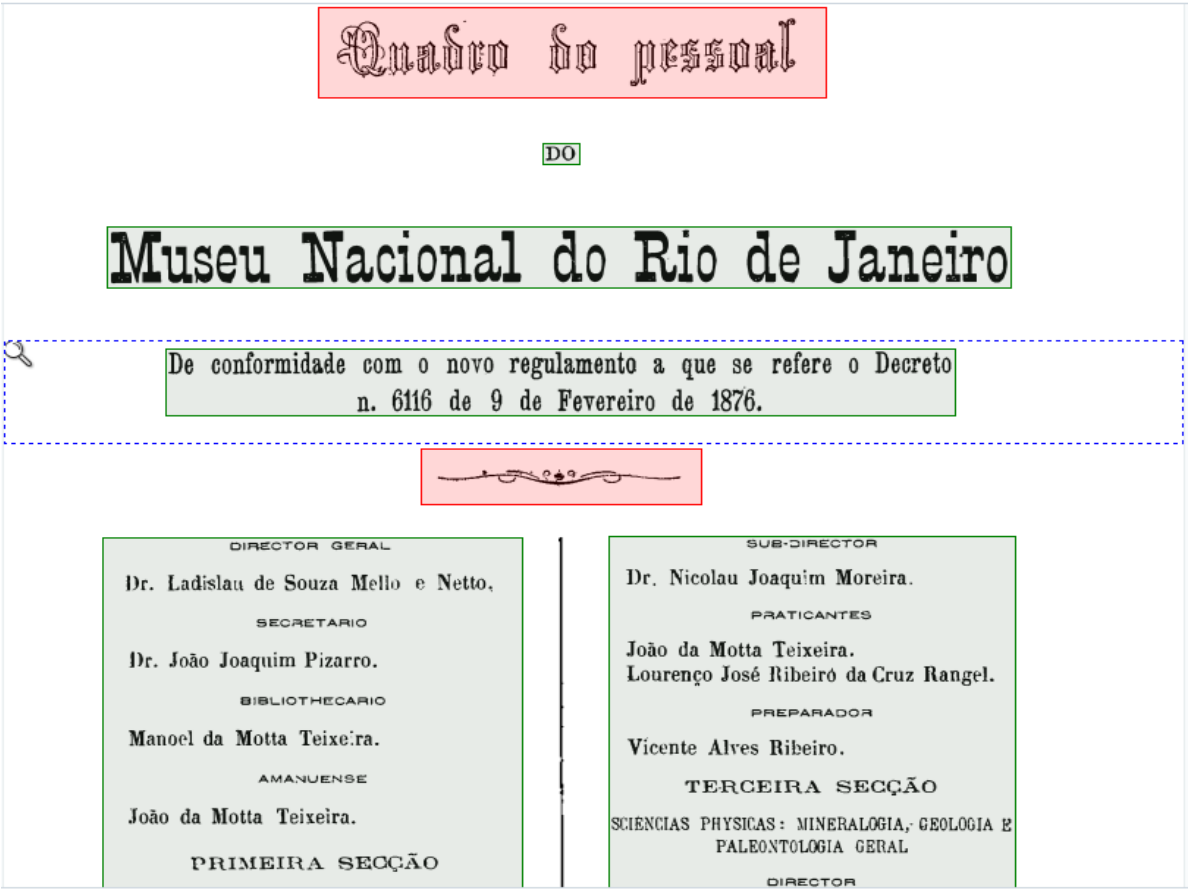
SCIENCIAS PHYSICAS: MINERALOGIA, GEOLOGIA E PALEONTOLOGIA GERAL

Fonte: Museu Nacional/Archivos do Museu Nacional/Autor.

4.3 OCR E DOCUMENTAÇÃO HISTÓRICA: O RUÍDO FAZ A DIFERENÇA

Dando sequência ao processo de digitalização do Volume 1 do *Archivos*, utilizo o *software Abbyy FineReader* para o pré-processamento das imagens, como etapa inicial para correção de irregularidades que possam extrapolar uma correta captura de informações. Na imagem a seguir (Figura 11), temos agora um recorte do documento com a aplicação de pré-processamento semiautomático, eliminando distorções, manchas e realinhando a imagem para um melhor reconhecimento de caracteres.

Figura 14: Recorte de documento destacando elementos reconhecidos e diferenciados de modo automático para processamento dos caracteres e demais objetos gráficos.



Fonte: Museu Nacional/Autor.

As seções da imagem, destacadas em vermelho, significam que o elemento em questão se trata de um gráfico ou imagem, já os elementos destacados em verde, são identificados como áreas de texto, e a depender do *software* implementado para realizar o processamento da imagem, a exatidão no reconhecimento de elementos e seções pode variar (a exemplo de texto em colunas).

Nem sempre o reconhecimento de caracteres se dá de forma eficiente, visto que o algoritmo responsável por esse processo precisa de uma boa diferenciação entre o caractere a ser reconhecido e o fundo da imagem, ou seja, para uma boa captura e reconhecimento, geralmente é aplicado um processo de alto contraste aos

elementos da página em comparação ao restante da imagem, como demonstrado nas figuras anteriores.

Para além do alto contraste como passo necessário para um bom reconhecimento do caractere, devemos levar em consideração outros problemas relacionados a variações linguísticas de um idioma, como por exemplo o português contemporâneo e o português arcaico, um vocabulário histórico ou mesmo uso de caracteres especiais e verbetes em latim, como bem explorado por Hauser (HAUSER, 2007).

Ainda assim, todo o processo de reconhecimento de caracteres deve ser mediado não só pelo computador, mas validado pelo usuário, verificado pelo fator humano ao final do processo. Apenas no reconhecimento dessa interdependência humano-máquina é que o processo de digitalização pode ser realizado com maior exatidão, senão em sua totalidade.

O "ruído" ao qual se refere esse momento da pesquisa se trata não apenas de informação não-indexada, que concretamente é todo e qualquer texto escaneado, mas também e principalmente dos caracteres reconhecidos erroneamente. Sendo de suma importância para a eficácia do processo de OCR, a identificação correta dos elementos textuais se dá por meio de "Aprendizagem de Máquina", um processo onde há uma inversão[26] de funções, onde o fator humano tem, literalmente, a palavra final ao definir se um caractere é ou não aquilo que o *software*, o algoritmo, reconheceu.

É com essa relação invertida, em que ensinamos a máquina o que ela deve corrigir, que partimos para outros processos necessários para a efetivação do OCR, como a construção de um banco de dados de padrões tipográficos.

4.3.1 *Reconhecimento de Padrões Tipográficos e a Construção de um Banco de Dados Linguístico*

Com essa exploração inicial do documento escaneado do periódico, nos deparamos com um número variado de tipografias que caracterizavam que tipo de

[26] Grifo do autor.

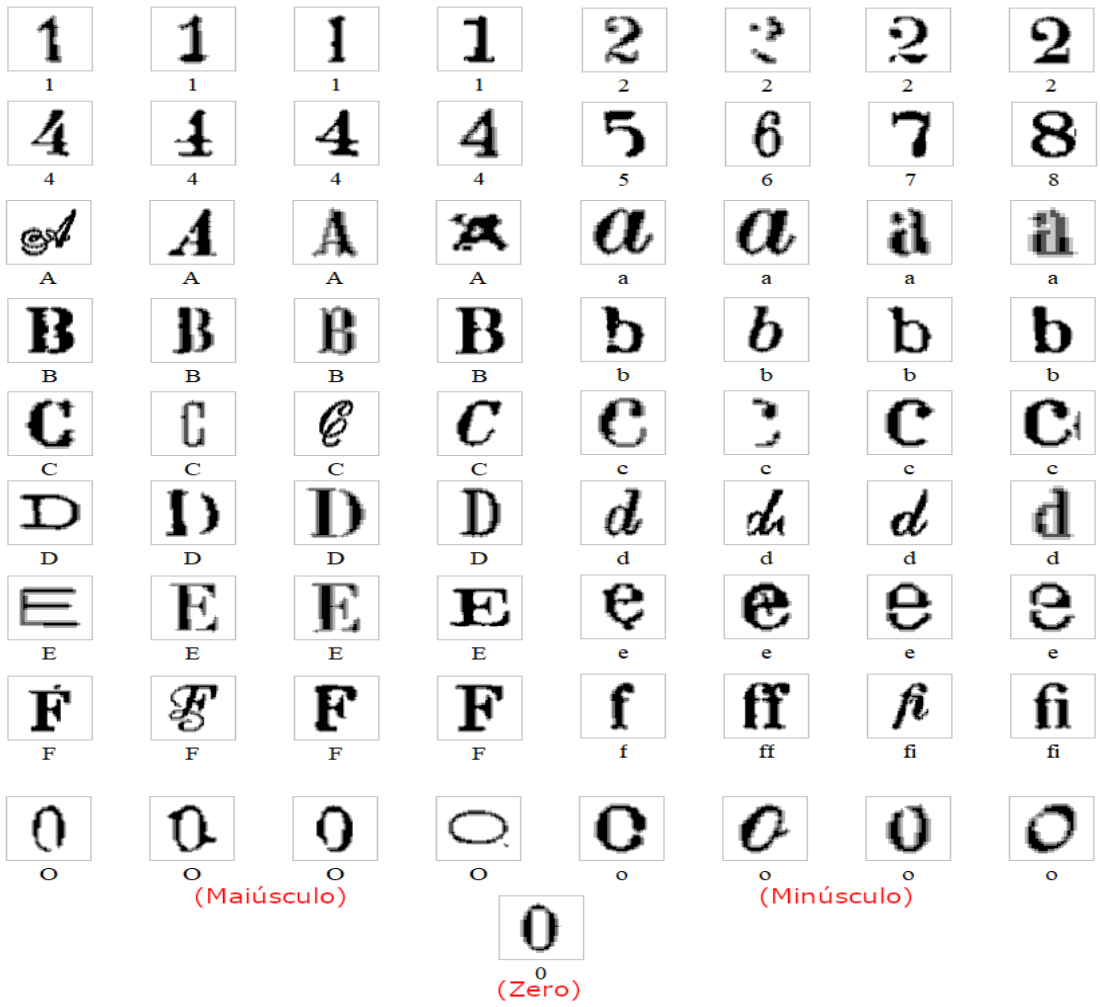
seção uma determinada forma textual designava, como também já pronunciava as dificuldades técnicas inerentes ao processo de reconhecimento uniforme dos caracteres do texto.

Nesse estudo de caso, o volume I do periódico do Museu Nacional apresenta entre três e cinco tipos de caracteres impressos no documento. Entre caracteres em maiúscula e minúscula, negrito, itálico e uma tipografia padrão para o corpo do texto, também encontramos elementos gráficos que servem de separadores de colunas e também dos rodapés, sendo esses últimos também passíveis de captura pelo *software* responsável pelo OCR.

Também se percebe um certo nível de dificuldade a partir de uma tipografia mais antiga (Fraktur, tipografia gótica), presentes geralmente em textos germânicos até meados do séc. XX, em que os caracteres apresentam angulações e decorações variadas que, nas experiências realizadas com o *Abbyy FineReader* (Figura 12), só obtivemos um resultado satisfatório a partir da Aprendizagem de Máquina, em que a interação entre o *software* e a confirmação dos padrões formaram um banco de dados de caracteres específicos.

A partir dos padrões elaborados na imagem abaixo, é possível perceber as variações no formato dos caracteres e estabelecer algumas noções de confiabilidade metodológica, como também mesmo sua ineficácia. Alguns casos são bastante notórios, como a identificação de similaridades entre numerais e letras (a exemplo do número 1 (um) e as letras *l* e *i)*, e mesmo entre letras maiúsculas e minúsculas (a exemplo da letra *o*, confundida as vezes também com o número *0)*.

Figura 15: Padrões tipográficos construídos a partir do Abbyy FineReader no processo de OCR do periódico Archivos do Museu Nacional v.I (1876).



Fonte: Abbyy FineReader/Autor.

O fato é que, para que haja um processo de transcrição do “analógico” para o digital, idêntico ao original, o processo de OCR deve reconhecer os caracteres como são apresentados tal qual, preservando inclusive erros de grafia. Em relação ao processo de construção de um banco de dados linguístico, deve-se atentar para as normas ortográficas ao longo da evolução da própria língua, recuperando a informação pretendida, ou seja, caso haja degradação do documento ao longo do tempo, e o mesmo se apresente com falhas de impressão, como sangramentos da tinta ou envelhecimento e ruptura de páginas, o contexto pode ou não fornecer a

informação ausente. E no caso do português identificado no periódico em análise, falado e escrito no século XIX, esse banco de dados teve de ser construído manualmente, de forma redundante, onde o próprio documento era a fonte de comparação para construir os padrões, já que o *software* responsável por todo esse processo possui codificado internamente apenas a língua portuguesa contemporânea pós-reforma.

Assim, para que o banco de dados seja estruturado corretamente, palavras como “farmácia” e “pharmacia” ainda que tenham o mesmo significado, são duas entradas diferentes no registro desse banco. Outros exemplos para o processo de identificação correta são as palavras com consoantes repetidas, como “admitir”, “annualmente", “effectuar”, “immanentes", assim como as grafias mais antigas citadas a pouco, como “actualmente”, “anthropologia”, “bibliotheca" e tantas outras.

Feitas as correções necessárias após o reconhecimento óptico dos caracteres, o banco de dados relacional é então estruturado, permitindo uma melhor exploração dos dados extraídos, à procura de pistas que sejam a base de uma análise textual que revele outras narrativas, que até então estavam limitadas por uma leitura linear do texto. Com a extração das informações, sua estruturação e o reconhecimento de termos específicos, é possível realizar uma modelagem de dados para potenciais narrativas até então não conhecidas. É o que dá origem ao termo “*Distant Reading”*, explorado em seguida.

4.3.2 *Transgressões da forma: close reading, distant reading e algo além.*

Chegamos até este momento da pesquisa para justificar todo uma cartografia processual no intuito de ousar transgredir a forma, e mesmo a ontologia do texto escrito. Estabeleceu-se um padrão de norma e de forma no que diz respeito a produção e publicação de conhecimento científico, qual seja, por meio de textos escritos. No entanto, as potencialidades apresentadas pelo avanço das tecnologias computacionais têm levado pesquisadores à experimentação com documentação escrita há mais de cinquenta anos.

Assim, a presente pesquisa também está se valendo dessas potencialidades para propor novas metodologias de análise documental, em que a transdução do texto escrito, tradicionalmente linear e unidimensional, pode coexistir com novas propostas ontológicas, em que o texto se torna modelável, rizomático e multidimensional.

Tomando de empréstimo um exercício estatístico de crítica literária feito por Franco Moretti (MORETTI, 2013) em sua obra "*Distant Reading"*, aplica-se técnica semelhante à análise do periódico "Archivos do Museu Nacional", objeto do nosso estudo de caso.

Como parte da laboriosa função de historiador, a leitura e pesquisa textual demanda uma grande parte do tempo disponível do sujeito na busca por informações relevantes. O documento textual tem em sua essência estruturas muito bem definidas, contanto com características que obedecem ao limite do meio empregado no processo da produção e publicação de conhecimento.

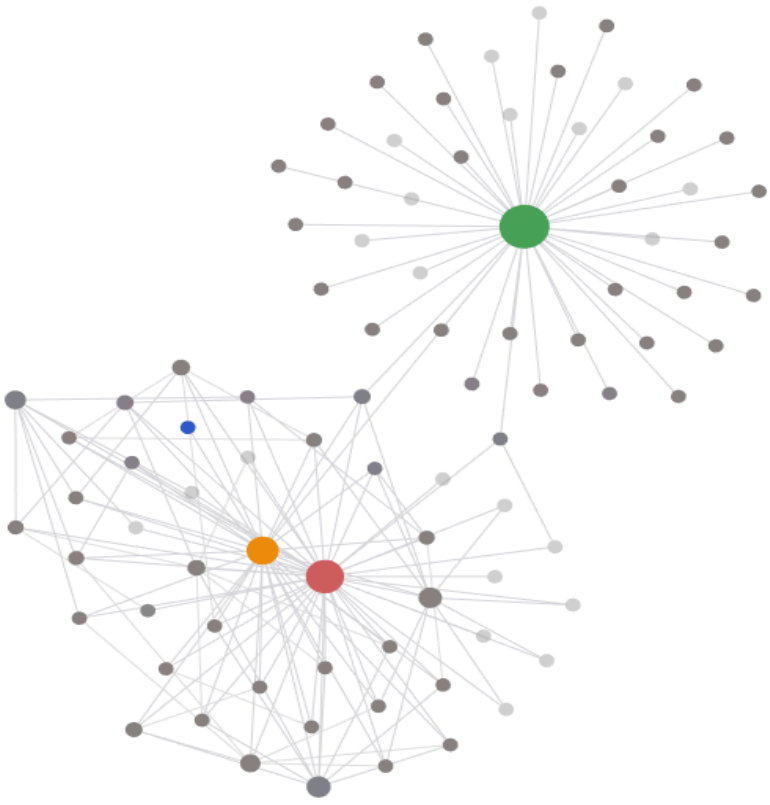
De modo geral, foi convencionado um território plano, sem muitas possibilidades criativas no que diz respeito a análise de seu conteúdo. Habitando o papel, a informação está enquadrada por margens, pontuações e regras que definem tamanho e espaçamento entre os caracteres. E quando se trata de um estudo científico, acadêmico, outras tantas regras vêm sendo aplicadas e seguidas à risca pela comunidade científica, ainda que novas tecnologias tenham surgido para expandir as possibilidades informacionais, que é o caso da presente pesquisa.

Assim, ao analisar detidamente o arquivo escaneado do periódico, foram extraídas as informações pertinentes que identificam determinadas entidades nomeadas e transcritas para a interface do ***Obsidian***, capaz de remodelar a informação ali representada, não apenas para dar nova forma ao conteúdo, mas pela busca de narrativas potencialmente ocultas no documento.

O modo de representação das informações extraídas dos documentos originais e implementadas em um *software* como o *Obsidian* é o que Moretti chamou de "d*istant reading"*, ou leitura distante, empregando uma análise estatística à literatura, e com o estudo do periódico do Museu Nacional, estamos experimentando também novos modos de representação e descoberta de conhecimento.

Na imagem seguinte (Figura 16), além da representação em grafos, há o recurso da classificação das entidades por cores, o que permite a identificação de agrupamentos específicos, uma "clusterização"[27]. Destacado em laranja está a entidade textual que representa a referência ao Museu Nacional, em vermelho a entidade que faz referência ao *Quadro de Pessoal do Museu Nacional*, e em verde a referência aos *Membros Correspondentes do Museu Nacional*, os demais *nós* dessa rede, representados na cor cinza são referências individuais a cada pessoa, lugar, instituição ou conceito que está expresso no documento em seu formato tradicional, escrito de forma linear.

Figura 16: Grafo de Entidades Nomeadas, informação extraída do vol. I de 1876 do periódico Archivos do Museu Nacioanal.



Fonte: Obsidian/Autor.

[27] Clusterização é um fenômeno inerente às dinâmicas estruturais das redes, representando um agrupamento de entidades com características semelhantes. Saber mais em COMMENDATORE, 2016.

Esse modo de representação de informações traz novas perspectivas não só para as análises textuais e classificação e delimitação de objetos de estudo, mas propõe uma nova ontologia documental, tendo como base uma metodologia computacional-imagética, capaz de acessar as dimensões do texto como um todo, sejam estas semânticas ou morfológicas.

À medida que as informações são transcritas para interfaces como a do *software Obsidian*, devemos considerar também as lacunas que um único documento apresenta. Por isso é importante a análise de um conjunto documental mais amplo (e também o uso de ferramentas diversas), que represente toda uma coleção de temas e referências, e que por mais extensas que sejam essas coleções, há uma dificuldade em estabelecer referências concretas entre documentos distintos, a menos que haja uma interface capaz de agregar toda essa informação, descortinando possíveis novas narrativas, ou mesmo relações já conhecidas, mas não apresentadas desse novo modo mais amplo.

Na experiência de transcrever para o *Obsidian* um artigo contido no periódico em sua forma integral, é possível constatar que a interação humano-máquina é constante, por meio de um reforço de aprendizagem e correção de informações. Principalmente quando se trata de verificar a eficácia do OCR, em que, ao ser executado por um único indivíduo, e mesmo com o auxílio de *software* especializado, alguns caracteres ainda são reconhecidos erroneamente, e o processo de transcrição requer um *feedback* também constante.

Outra correção importante é o referenciamento correto das entidades nomeadas extraídas do documento. Haja vista que métodos para citação e escrita de artigos não obedeciam a um padrão institucional no séc. XIX e assim, referências como “Burton” (MUSEU NACIONAL. Archivos. v. I. p.15), precisaram ser analisadas a partir de uma fonte externa à coleção que é foco do estudo de caso[28].

28 Também há uma referência que chamou bastante atenção no que diz respeito a um indivíduo sem sobrenome, conhecido e citado apenas como Pitanga, o que leva a crer que se trata de um escravo, participando de uma expedição científica, como guia local ou rastreador.

4.4 CARTOGRAFIAS DO MUSEU NACIONAL: O ESPAÇO DE FLUXOS DAS CIÊNCIAS NO BRASIL DO SÉC. XIX SOB A ÓTICA DAS HUMANIDADES DIGITAIS NO SÉC. XXI

Ao considerar toda a caminhada da pós graduação nesses últimos cinco anos, com as experimentações em análise documental, utilizando diversos *softwares* especializados na construção de novas ontologias documentais, não resta dúvida que o que se entende por *cartografia processual* é uma metodologia interdisciplinar que vem traduzindo de forma ampla a metáfora de rede que se estabeleceu como pilar argumentativo dessa tese.

Enquanto processo fundamental da produção de saberes, uma ontologia não-linear da representação da informação, de forma visual e auxiliada por métodos e técnicas computacionais, se efetivou na transgressão e transdução do conhecimento histórico e científico das fontes primárias analisadas nesse capítulo experimental. O mergulho na documentação do Museu Nacional teve não só um caráter que objetivou a inovação, mas um chamado emergente para novas práticas de produção de saberes enquanto política científica de preservação documental num contexto político de constante tensão, em que as verdades factuais dos documentos estão a mercê do apagamento, que sejam por falta de acesso, ou literalmente consumidos pelo fogo.

Talvez por coincidência, os primeiros volumes do periódico *Archivos do Museu Nacional* foram elaborados num contexto de muitas transformações no cenário político e, como atualmente, são parte da preocupação dessa pesquisa para compreender as relações que se estabelecem entre sujeitos, instituições e ideias e a classe política e o cidadão comum, tanto na elaboração de políticas públicas em ciências como para o reconhecimento dos discursos científicos no imaginário do povo.

Compreendendo o contexto do lugar dos documentos do periódico na pesquisa, retoma-se aqui o que Manuel Castells estabeleceu como um *espaço de fluxos* (CASTELLS, 1999, p. 467), enquanto um lugar central para a circulação não só de pessoas, como de informações e ações políticas em ciências. E o desafio tem sido

colocar em nova perspectiva, principalmente sob a ótica das Humanidades Digitais, a dinâmica das relações que se estabeleceram e se eternizaram na documentação sobre a História das Ciências no Brasil no contexto do séc. XIX, a partir de métodos do séc. XXI.

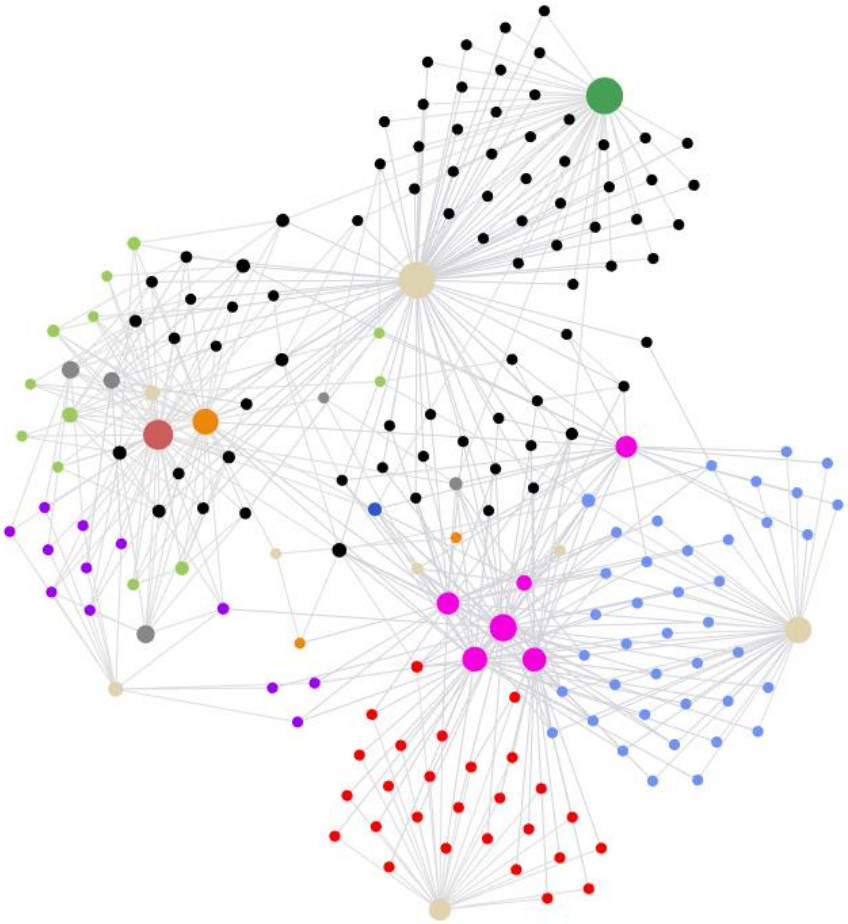
Utilizar um conceito como *espaço de fluxo*, elaborado a partir de um contexto informacional e tecnológico próprio do séc. XXI, e aplicá-lo a um espaço-tempo passado requer adaptações metafóricas que permitam a identificação de características em comum através do tempo. Segundo Castells, o espaço de fluxo se constitui a partir de 3 elementos:

a Um meio de fluxo adequado.
b Elementos que compõem um fluxo e que se "atravessam" por meio dele.
c Os "nós" que organizam, controlam e condicionam a circulação dos elementos no fluxo.

Assim, tomando o Museu Nacional como epicentro das relações políticas e científicas no séc. XIX, tem-se a identificação do primeiro elemento necessário para o reconhecimento do espaço de fluxos. Em seguida, considerando os sujeitos que transitavam e que, de algum modo, se comunicavam com esse espaço, tem-se os elementos compondo o fluxo. E por fim, o que é chamado de "nós" nessa metáfora de rede, são as relações estabelecidas entre os sujeitos, instituições e ideias que se atravessam a partir de uma análise egocentrada no Museu Nacional e o periódico *Archivos do Museu Nacional*, em que consta o registro escrito desses atravessamentos, possibilitando assim sua transcrição por meio de metodologias computacionais[29].

[29] MAIA, D. 2017

Figura 17: Cartografia das relações científicas no periódico Archivos do Museu Nacional



Fonte: Obsidian/Autor

A partir de uma ontologia imagética (Figura 14), resultado de um processo de transgressão e transdução do arquivo escaneado, pude elaborar algumas conexões entre as entidades nomeadas, classificadas em oito categorias de informação.

Como pode ser visto abaixo (Figura 15), as categorias de Entidades Nomeadas fazem referência a informações específicas contidas nas publicações do periódico, em que é possível classificá-las de acordo com parâmetros determinados pelo tema colateral da pesquisa, qual seja, a formação de uma rede epistêmica complexa, em que campos científicos e conceitos, se relacionam direta ou indiretamente com

instituições, pessoas, cargos e lugares enquanto componentes de um espaço de fluxos que representam as relações e atividades científicas estabelecidas no contexto da escrita dos artigos do periódico.

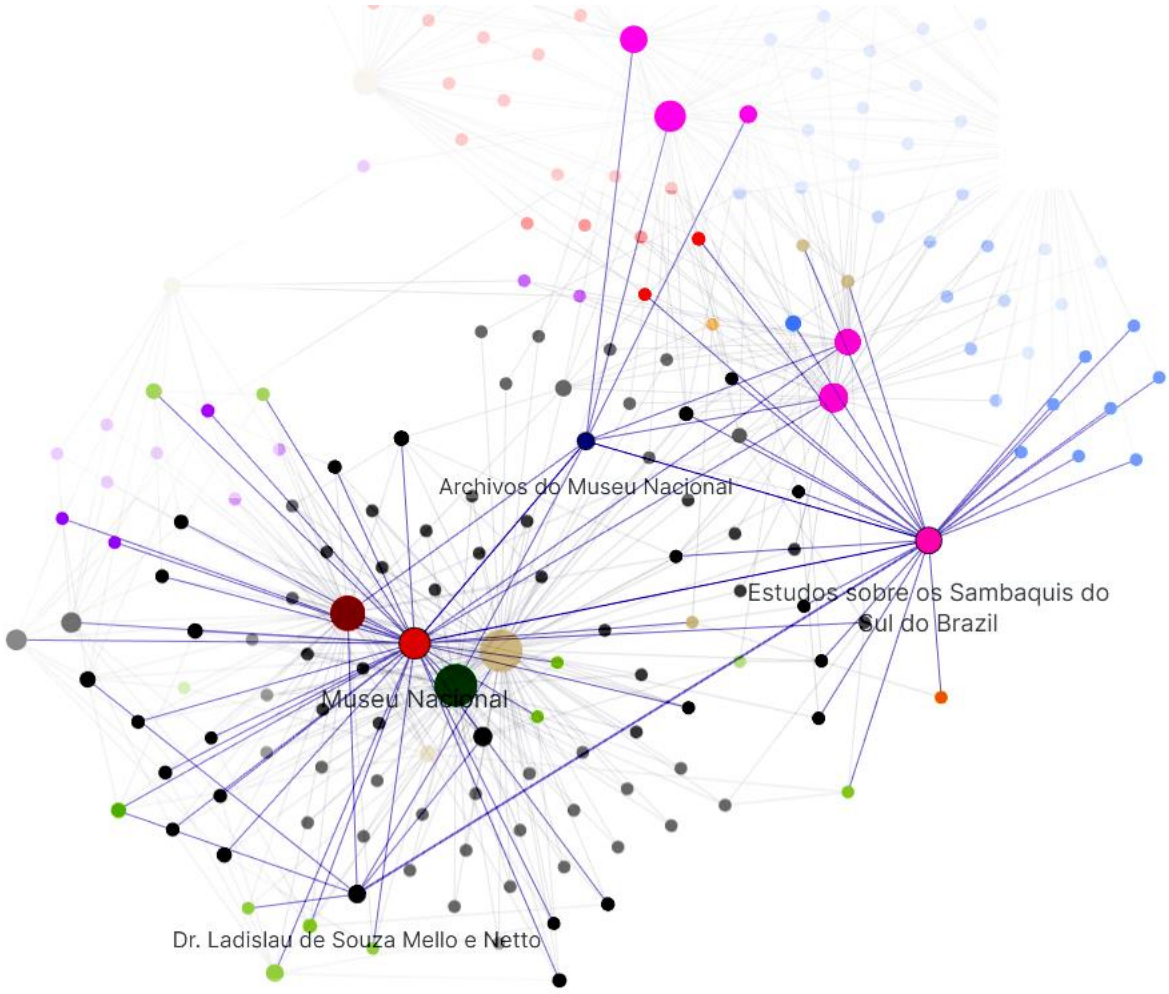
Figura 18: Interface do Obsidian com Hashtags e Categorias de Entidades Nomeadas

Fonte: Obsidian/Autor

Ainda tomando como referência a Figura 14, é possível compreender a importância da classificação das entidades nomeadas como um processo de formação e identificação de *clusters* ou agrupamentos de informações com características em comum, sendo de notável importância também o uso de cores como atribuição de características distintas entre as entidades.

É possível visualizar na interface do *Obsidian* as microrrelações estabelecidas a partir de qualquer entidade nomeada, ou seja, qualquer nó da rede que ali se constitui, para melhor identificação das conexões.

Figura 19: Microrrelações entre quatro entidades nomeadas.



Fonte: Archivos do Museu Nacional/Obsidian/Autor

Exemplo dessas microrrelações são mostradas acima (Figura 19), em que se identificam 4 entidades nomeadas formando uma rede, ao relacionar a instituição Museu Nacional, seu periódico *Archivos do Museu Nacional*, a atuação de seu diretor geral em 1876, Ladislau Netto, e alguns temas produzidos nesse primeiro volume, como a questão dos Sambaquis no Brasil. Os demais nós da rede, não identificados por uma limitação da interface do *software*, representam outros agrupamentos, como os membros correspondentes (em preto) e lugares mencionados nos artigos publicados no periódico (em azul), sejam estes locais de pesquisa ou origem de outras instituições e pesquisadores que se correspondiam ou avaliavam pesquisas para o

Museu Nacional, o que fortalece a tese de que o mesmo era um epicentro dinâmico enquanto espaço de ciências no Brasil.

> Em conjunto com a documentação do Museu Nacional, a bibliografia sobre Dom Pedro II revelou seu papel essencial, porém não estrutural, na rede de agenciamentos. Revelou por sua vez, que uma rede constituída de diversos personagens atuantes possui múltiplos centros de agenciamento, e no caso do Museu Nacional, este ocupava um papel central na estrutura das relações institucionais, visto que congregava um grande número de colaboradores nas atividades científicas e, por meio da publicação de seu periódico, realizava permutas com instituições nacionais e estrangeiras, fazendo com que o conhecimento fluísse. (MAIA, 2017, p. 81).

Já numa segunda abordagem experimental, o segundo volume dos “Archivos do Museu Nacional” traz uma bibliografia de permutas realizadas entre instituições nacionais e internacionais, para onde foram enviados exemplares do periódico, e isso dá uma pista da extensão das redes de relações estabelecidas pela instituição, como também permite a identificação de similaridades nas políticas científicas adotadas pelo desenvolvimento das mesmas no cenário internacional.

Figura 20: Cartografia das permutas dos Archivos do Museu Nacional.

Fonte: Obsidian/Autor.

Na imagem anterior (Figura 17), está representado parte do grafo em que é possível identificar as permutas relacionadas ao periódico *Archivos do Museu Nacional*, e também parte das relações estabelecidas com instituições internacionais, sendo algumas em destaque como os *Anales de Museo Publico* de Buenos Aires, em que vemos conectado à rede Germán Conrad Burmeister, naturalista prussiano, diretor do mesmo museu.

A partir da identificação da função de diretor do museu público de Buenos Aires, segui a pista das publicações que se atravessam pelos nomes de Burmeister e Ladislau Netto entre outros, na perspectiva de construção de uma rede epistêmica que é expressa de maneira potencialmente ampla nos grafos da interface digital do *Obsidian*.

Na busca por mais informações sobre Burmeister, por exemplo, chegamos ao artigo "*A mesma fé e o mesmo empenho em suas missões científicas e civilizadoras: os museus brasileiros e argentinos do século XIX"* (LOPES, 2001), este publicado pela Revista Brasileira de História e indexado pela Scielo. Ao analisar as notas bibliográficas do artigo, outras redes epistêmicas acabam por se formar, estabelecendo uma ponte entre os escritos do século XIX e os do século XXI, por sua vez revisitados e modelados a partir da proposta dessa experimentação metodológica.

É senso comum que a disponibilidade remota de fontes primárias seja bastante restrita em pesquisa histórica, e não com menos dificuldades o acesso a acervos internacionais acaba por limitar e muito pesquisas bibliográficas, justificando, portanto, que os projetos de digitalização de acervos seja cada vez mais encorajado pelas instituições que lidam com memória documental. Na busca por mais informações sobre o Museu Publico de Buenos Aires, recorremos ao portal "*Internet Archive"* para localizar alguma referência ao banco de digitalizações da instituição, não só revelando outros volumes dos "Anales do Museu Publico", como também alcançando a "*Biodiversity Heritage Library"*, outro portal de memória digital institucional dedicado as ciências e seus arquivos digitalizados.

Desse modo a construção da rede epistêmica tem seu potencial transformado, dando acesso a pesquisadores estrangeiros que estejam em busca de documentação que complemente suas pesquisas sem necessariamente terem de se deslocar ao local de origem das instituições pesquisadas, e isso foi válido para o século XIX, por meio de periódicos permutados, como é válido hoje com as publicações e armazenamento digital de artigos científicos, com a diferença da velocidade de transmissão das informações e conhecimentos e das interfaces utilizadas.

Assim, de modo análogo a um rizoma, desenvolve-se o potencial para a formação e descoberta de uma rede epistêmica colateral de informações e conhecimentos. Em que a representação visual traz, de modo amplo, toda a cartografia processual oriunda da análise de um conjunto documental muito além daquele planejado inicialmente como recorte espaço-temporal de uma pesquisa.

E isso só se torna possível quando consideramos o potencial agregador dos repositórios digitais disponibilizando acervos documentais, digitalizados ou não, mas

que permitam o acesso integral a informação, não apenas as referências catalogadas, pois é no corpo do texto, no território das letras, que se constitui tal cartografia.

CONSIDERAÇÕES FINAIS (A FUGA DO GATO DE SCHRÖDINGER)

Fronteiras, cercas, limites, demarcações, linhas, enquadramento, epistemologias... Padrões. O peso da adjetivação do pensar foi medido ao acender das luzes da Modernidade e com isso a responsabilidade de procurar a verdade dentro dos moldes da percepção racionalizada se instalou na mente daqueles que têm tido o papel de desenvolver o conhecimento científico, registrar a história e garantir as estruturas da realidade no agir humano para com o mundo.

Inadvertidamente, ou por ironia de forças que se auto-organizam, a complexidade da relação do ser humano com a natureza tem nos apresentado as consequências dessa fragmentação interpretativa da realidade. Atualmente, novos sujeitos têm sido inseridos na nossa dimensão social para avançarmos na compreensão de mundo: artefatos sociotécnicos produzidos pela mente humana, alimentados com nossas linguagens e nossos códigos, mas com maior poder de processamento de informações, assumiram parte dessa tarefa de interpretação.

É no reconhecimento desses novos sujeitos que a presente investigação sobre Humanidades Digitais trouxe resultados complexos sobre metodologias transdisciplinares como uma transgressão nos diálogos entre os campos científicos.

O trabalho de pesquisa acadêmica é uma atividade genuinamente individual, em que os processos cognitivos se desenvolvem isoladamente, ainda que tenhamos dados e estudos que nos acompanham nas nossas reflexões. Enquanto área emergente e transdisciplinar, as Humanidades Digitais ainda estão estabelecendo suas fundações, e como tal, caminhos heterogêneos percorridos por indivíduos têm sido feitos, mas em direção a um ethos de colaboração, da identificação da prática coletiva como base ontológica tanto nas práticas como nas teorias que integram os estudos em Humanidades Digitais.

A ideia de “laboratório”, por exemplo, tomando de empréstimo a ideia latouriana, incide sobre as humanidades de uma maneira diferente, haja vista que as

práticas em Humanidades Digitais não se dão exatamente dentro de um prédio, hermeticamente preparado para analisar os objetos a serem estudados com um olhar microscópico. O laboratório das Humanidades Digitais talvez se constitua na observação e interação entre sujeitos e seus objetos num contexto global, cibernético ou não, sem condições ideais pré-estabelecidas. E no tocante a esta pesquisa, acredito que o próprio contexto global de pandemia foi o laboratório que se apresentou e colocou à prova os limites e potencialidades dessa interação globalizada, onde muitos de nós pesquisadores precisaram se isolar e lidar com nossas reflexões, nossas fontes e objetos sem os processos de interação social mais tradicionais.

As limitações da narrativa textual observadas na tese refletem um conflito ontológico entre o linear e o não-linear, o que foi proposto na pesquisa. A dificuldade de leitura de um texto estático revela que os processos cognitivos naturais da mente humana necessitam de um dinamismo e de uma capilaridade na apreensão das relações entre ideias. O texto estático, longilíneo, parece não oferecer isso, ao contrário do que um quadro geral que apresenta todo o dinamismo de uma única vez, permitindo ao leitor uma livre cartografia na identificação e apreensão do que se lê.

Ainda assim, foi possível constatar um novo fazer historiográfico e científico mediado com técnicas e ferramental computacional facilitando, e mesmo aprimorando, a lida com documentação textual tanto o objeto físico como o digital.

Como esperado, a hipótese de que a adoção de métodos e técnicas computacionais traria celeridade aos processos tradicionais de pesquisa, verificação e análise documental mostrou que nos encontramos em um potencial de alto nível tecnológico para as humanidades, resolvendo antigas questões das atividades historiográficas frente a um grande volume de documentos.

Em conjunto com tais potencialidades sociotécnicas, surgem novas questões ainda mais complexas, haja vista a rápida e vertiginosa inserção de um novo elemento na relação da sociedade com a produção, disseminação e acesso ao conhecimento.

Com a aceleração das pesquisas em Inteligência Artificial e as possibilidades de melhoramento de mídias, como imagens, sons, vídeos e também textos, a preocupação com a falsificação atingiu um patamar perigoso em que os danos já

podem ser percebidos, não só no imaginário popular, mas refletindo também nas leis e protocolos de segurança e privacidade de diversos países.

O desenvolvimento de tecnologias conhecidas como *deepfakes* tem trazido à tona preocupações com a disseminação de discursos enganosos, muitas vezes carregados com narrativas políticas de perseguição a minorias e opositores partidários[30], mas que também tem potencial para alterar a percepção sobre a memória histórica coletiva, onde os meios de comunicação são utilizados para alcançarem principalmente os menos atentos na diferenciação do fato real e do fabricado.

Em termos de investigação de base documental e de artefatos, devemos lembrar que as falsificações não são algo novo. Desde falsificações de pinturas, esculturas e selos governamentais, há registros já na antiguidade sobre esse tipo de atividade. A diferença é que hoje lidamos com outras dimensões de manipulação factual, em que os *bits* de informação são os verdadeiros alvos das falsificações, alterando as microestruturas da informação e não o artefato ou o documento em si. É aí que entram os recursos de criptografia avançada, atribuições de valores aleatórios como garantia de uma identidade digital única para cada tipo de objeto.

No que diz respeito a documentação histórica e a segurança do dado histórico contido no texto digital, ainda é incerto que tipos de protocolos podem ser construídos para salvaguardar os novos tipos de mídia, levando em consideração não só o maquinário de armazenamento e sua manutenção perpétua, como também as barreiras para impedir invasões e a potencial exclusão dos bancos de dados de documentos e coleções. Aparentemente ainda estamos num longo processo de transição informacional, um movimento acelerado de transformação de antigas mídias físicas para o meio digital, incluindo a construção de modelos 3D de artefatos

[30] Para exemplificar, lembremos do documentário, indicado ao Oscar, "Democracia em Vertigem", dirigido pela Petra Costa, em que uma polêmica se instalou quando questionada sobre a adulteração digital de uma fotografia histórica que mostrava a ausência da presença de uma arma, plantada na cena do assassinato de dois dirigentes do Partido Comunista (PC do B).

arqueológicos, como escaneamento de sarcófagos sem a manipulação direta de seu conteúdo, e mesmo de livros em avançado estado de deterioração.

Os avanços com a experimentação teórico-metodológica para com o conjunto de textos históricos nos permitiram elaborar novos protocolos de preservação, publicização e acesso ao conhecimento, onde já havia políticas de digitalização documental em diversas esferas institucionais, sejam estas da comunidade científica ou do próprio poder público. Com a experimentação, foi possível propor como novos processos digitais podem ser adotados, levando em consideração as rápidas transformações e desenvolvimento de tecnologias no campo da Aprendizagem de Máquina e Inteligência Artificial. Foi possível também identificar lacunas e demandas por uma formação transdisciplinar de áreas tradicionais do conhecimento, especialmente a História, quando confrontamos métodos tradicionais das Humanidades e novos métodos da Computação.

Daí vem a escolha pelo termo “emergência”, que compõe o subtítulo da tese. Não é apenas a expressão de um campo em rápido desenvolvimento, mas também uma resposta imediata às questões contemporâneas que se apresentam na resolução de desafios que vão muito além de questões epistemológicas nas ciências, como o enfrentamento da escalada das *Fake News* como projeto político fascista e antiecológico em escala global. Propor novos protocolos metodológicos de preservação documental, com transparência e acesso universal, é uma medida garantidora para o desenvolvimento social e científico.

Isso abre caminho para novas questões metodológicas e interações entre disciplinas e cientistas de campos distintos, algo que foi explorado nesta pesquisa com relativo sucesso. Mesmo com 50 anos de desenvolvimento, o campo das Humanidades Digitais parece renovar de forma acelerada a atmosfera de atuação das Humanidades como um todo, e no contexto brasileiro, a emergência do campo carrega de fato o duplo sentido para assegurar a preservação e a inovação na pesquisa científica.

Espera-se que, com a conclusão desta pesquisa, tais questões complexas ganhem espaço em agendas que abranjam aspectos psicossociais, econômicos e

políticos da atividade científica no Brasil e do papel dos sujeitos das ciências, incluindo o que as instituições não conseguem quantificar e avaliar a partir de critérios tradicionais, sejam estes relacionados às normatizações de escrita científica e processo criativo, sejam na avaliação dos campos de conhecimento, suas produções intelectuais e as relações econômicas resultantes de uma fragmentação epistemológica herdada da Ciência Moderna.

Como um processo transgressor, abordagens sistêmicas que consideram múltiplas dimensões de agenciamento entre sujeitos, instituições, artefatos e ideias abrem novos caminhos para reintegrar os saberes dando-lhes novos significados. E no momento atual, em que as ciências sofrem ataques profundos, transformando o imaginário social, colocando em xeque questões científicas que afetam diretamente a vida em sociedade, movimentos de transgressão nas ciências são um dever político para o desenvolvimento humano e global.

A falsa dicotomia entre direitos autorais científicos e o acesso à informação vai além de questões puramente econômicas. É notório que os grandes atores institucionais das ciências deixaram de lado o valor universal do conhecimento, construindo barreiras intangíveis de salvaguarda documental, reforçando a desigualdade social por meio do controle do desenvolvimento científico e tecnológico. Em contrapartida, parte da comunidade científica vem adotando uma atitude ética e combativa, abraçando o conceito de ciência aberta para equalizar essa balança de desigualdades.

No âmbito institucional, universidades e centros de pesquisa têm adotado a indexação de suas pesquisas em plataformas de acesso aberto, permitindo a pesquisadores de todo o mundo a consulta integral de documentação para complementação de seus trabalhos. No aspecto combativo, a comunidade científica tem contornado as barreiras intangíveis com táticas de guerrilha digital, haja vista a iniciativa da Alexandra Elbakian com o desenvolvimento do *Sci-Hub*[31], e figuras como

[31] O modo de funcionamento do site Sci-Hub se dá pela quebra da chamada *Pay-Wall*, dando acesso direto e gratuito aos artigos científicos codificados com o DOI.

o Aaron Swartz[32], que chegou a invadir servidores da sua universidade, roubando milhões de artigos científicos e disponibilizando ao público. Iniciativas como essas inspiraram e permitiram a realização da presente tese de doutorado.

Por fim, não adianta pensarmos em inovação sociotécnica sem assegurarmos os protocolos de preservação de toda infraestrutura, incluindo o armazenamento dos documentos físicos, as políticas científicas de salvaguarda institucional, a formação técnica e intelectual dos sujeitos nos espaços de ciências, como também a promoção da importância da memória histórica e científica para a sociedade.

.

[32] A história de Aaron Swartz foi contada no documentário "O Garoto da Internet" (*The Internet's Own Boy*), apresentado na disciplina "Ciências Quadro à Quadro", durante o Doutorado. (ANEXO 3)

REFERÊNCIAS

ABBYY. **Abbyy FineReader**. Página inicial. Versão 15. [S.I.], 1989. Disponível em: <https://www.abbyy.com/>. Acesso em: 22 Jan. 2017.

ALONSO, Angela. **Ideias em Movimento**: a Geração 1870 na crise do Brasil-Império. São Paulo: Paz e Terra, 2002.

ARCHIVOS DO MUSEU NACIONAL. Museu Nacional: Imprensa Industrial, v. 1, 1876.

ARCHIVOS DO MUSEU NACIONAL. Museu Nacional: Imprensa Industrial, v. 2, 1877.

Avanços Tecnológicos no Cinema. TV Gazeta. Disponível em: <https://www.tvgazeta.com.br/videos/avancos-tecnologicos-no-cinema-fevereiro-2020/>. Acesso em: 22 de fevereiro de 2021.

BERGSON, H. O pensamento e o movente – Introdução. *Bergson*. São Paulo: Abril Cultural (Col. Os Pensadores), 1979 (texto originalmente publicado em 1934).

BARABÁSI, Albert-László. ***Linked****: The New Science of Networks*. Cambridge, USA: Perseus Publishing, 2002.

_____.et al. ***The Structure and Dynamics of Networks***. *New Jersey: Princeton University Press, 2006.*

BATESON, Gregory. ***Steps to an Ecology of Mind****: collected essays in Anthropology, Psychiatry, Evolution and Epistemology*. 1972. San Francisco, USA: Chandler Pub. Co., 1972.

_____. ***Mind and Nature****: A Necessary Unity*. New York, USA: E. P. Dutton, 1979.

BARROS, José D'Assunção. **O Campo da História**: especialidades e abordagens. Petrópolis, RJ: Vozes, 2004.

BERRY, David M. ***Understanding Digital Humanities***. Basingstoke, UK: Palgrave Macmillan, 2012.

BRACHMAN, Ronald J. ***Readings in Knowledge Representation***. California, USA: Morgan Kaufmann Publishers Inc., 1985.

BRANDES, Ulrik; ERLEBACH, Thomas. (orgs.) ***Network Analysis****: methodological foundations. Berlin: Springer-Verlag Berlin Heidelberg, 2005.*

BULLOUGH, Vern L. *The Computer and the Historian: some tentative beginnings.* ***Computers and the Humanities****. Volume 1. Issue 3. 1967*. p. 61-64. Disponível em: <https://sci-hub.st/10.1007/BF00119888>. Acesso em: 15 Jul 2020.

CARDOSO, Ciro F; VAINFAS, Ronaldo. **Domínios da História**: ensaios de Teoria e Metodologia. 5.ed. Rio de Janeiro: Campus, 1997.

_____. **Novos Domínios da História**. Rio de Janeiro: Elsevier, 2012.

CASTELLS, Manuel. **A Sociedade em Rede**. Tradução de Roneide Venancio Majer. Atualização para 6. ed. de Jussara Simões. (A Era da Informação: economia, sociedade e cultura. vol. 1). São Paulo: Paz e Terra, 1999.

COMMENDATORE, Pasquale. *et al*. ***Complex Networks and Dynamics****: Social and Economic Interactions. (Lecture Notes in Economics and Mathematical Systems) Switzerland: Springer International Publishing, 2016*.

CHEN, Chaomei. ***Mapping Scientific Frontiers***. *2. ed. Pennsylvania: Springer, 2013.*

CHERIET, Mohamed. *et al*. ***Character Recognition Systems****: a Guide for Students and Practitioners. Hoboken, US: John Wiley & Sons. 2007*.

CHERVEN, Ken. ***Mastering Gephi Network Visualization****. Birmingham, UK: Packt Publishing, 2015*.

CHRISTAKIS, Nicholas A. FOWLER, James H. **O Poder das Conexões**: a importância do networking e como ele molda nossas vidas. Tradução Edson Furmanikiewicz. Rio de Janeiro: Elsevier, 2010.

DANTAS, Regina Maria Macedo Costa. **A Casa do Imperador: Do Paço de São Cristóvão ao Museu Nacional**. Rio de Janeiro: 2007. Dissertação (Mestrado em Memória Social) – Universidade Federal do Estado do Rio de Janeiro.

_____. **Casa Inca ou Pavilhão da Amazônia? A participação do Museu Nacional na Exposição Universal Internacional de 1889 em Paris**. Rio de Janeiro: 2012. Tese (Doutorado em História das Ciências e das Técnicas e Epistemologia) – Universidade Federal do Rio de Janeiro.

DANTES, Maria Amélia M. et al.(orgs.). **A Ciência nas Relações Brasil-França (1850-1950)**. São Paulo: FAPESP, 1996.

_____. (org.) **Espaços de Ciência no Brasil**: 1800-1930. Rio de Janeiro: Editora Fiocruz, 2001.

DELEUZE, Gilles. GUATTARI, Felix. **Mil Platôs** - Capitalismo e Esquizofrenia v. 1. Tradução de Aurélio Guerra Neto e Celia Pinto Costa. São Paulo: Editora 34, 1995.

DOBSON, James E. ***Critical Digital Humanities****: the Search for a Methodology. Springfield, USA: University of Illinois Press, 2019.*

DYNALIST INC. ***Obsidian***. Página inicial. Versão 0.12.12. [S.I.], 2020. Disponível em: <https://*Obsidian*.md>. Acesso em: 22 Jan. 2021

FLANDERS, Julia. JANNIDIS, Fotis. **The Shape of Data in the Digital Humanities**: Modeling Texts and Text-based Resources. New York, USA: Routledge Taylor & Francis Group, 2019.

FLUSSER, Vilém. **Into the Universe of Technical Images**. Translated by Nancy Ann Roth. Minneapolis, USA: University of Minnesota Press, 2011.

FREUD, S. “Recomendações aos médicos que exercem a Psicanálise”. **Obras Completas de Sigmund Freud**. v. XII. Rio de Janeiro: Imago, 1969. (texto originalmente publicado em 1912).

GITELMAN, Lisa. **Paper Knowledge**: Toward a Media History od Documents. [S.l]. Duke University Press, 2014.

GRANOVETTER, Mark S. The Strength of Weak Ties. **American Journal of Sociology**. Volume 78. Issue 6. 1973. p. 1360-1380. Disponível em: <https://sci-hub.st/10.1086/225469>. Acesso em: 15 Nov. 2019.

HARARY, Frank. ***Graph Theory***. *Massachusetts: Addison-Wesley Publishing Co., 1969.*

HAUSER, Andreas W. **OCR Postcorrection of Historical Texts**. Magister Artium (M.A). Centrum für Informations- und Sprachverarbeitung (CIS). Ludwig-Maximilians-Universität München. 2007.

HAYLES, Katherine. How We Think: Transforming Power and Digital Technologies. In: BERRY, David M. (ed.). **Understanding the Digital Humanities**. Basingstroke, UK: Palgrave Macmillan, 2012.

HUITFELDT, Claus. Multi-dimensional texts in a One-dimensional medium. **Computers and the Humanities**. Volume 28. Issues 4-5. 1994. p. 235-241. Disponível em: <https://sci-hub.st/10.1007/BF01830270>. Acesso em: 22 jun. 2019.

IBICT. **LARHUD**, 2021. Disponível em : http://www.larhud.ibict.br/. Acesso em 22 fev. 2021.

INTERNET ARCHIVE. **Internet Archive**. Página inicial. [S.I.], 1996. Disponível em: <https://archive.org>. Acesso em: 14 Jan 2021.

KASTRUP, Virginia.; PASSOS, Eduardo.; ESCÓSSIA, Liliana da. **Pistas do Método da Cartografia**: pesquisa-intervenção e produção de subjetividade. v.1. Porto Alegre: Sulina, 2010.

_____. **Pistas do Método da Cartografia**: pesquisa-intervenção e produção de subjetividade. v.2. Porto Alegre: Sulina, 2016.

KHOKHAR, D. Eurosis Web Mapping Study. In: KHOKHAR, D. ***Gephi* Cookbook**. Birmingham, UK: Packt Publishing, 2015.

KOVALCHENKO, Ivan D. **The Russian Serf Peasantry in the First Half of the 19th Century**. Moscow: Izdatel'stvo Moskovskogo Universiteta, 1967.

KUHN, Thomas S. **The Structure of Scientific Revolutions**. 3.ed. Chicago, USA: The University of Chicago Press. 1962.

LACERDA, J. B. de. **Congresso Universal das Raças em Londres 1911**. Rio de Janeiro: Museu Nacional, 1912.

_____. **Fastos do Museu Nacional do Rio de Janeiro**. Rio de Janeiro: Imprensa Nacional, 1905.

LADURIE, Emmanuel Le Roy. **The Territory of the Historian**. Translated from the French by Ben Reynolds and Siân Reynolds. Brighton, UK: Edward Everett Root Publishers, 2017.

LEHNERT, Herbert.; RUECKING, Frederick.; PORTER, Robert. The Thomas Mann Project. **Computers and the Humanities**. Volume 1, Issue 3. 1967. p. 65-71.

LEVITT, James. H.; LABRARRE, Claude E. Building a Data File from Historical Archives. **Computers and the Humanities**. Volume 9, Issue 2. 1975. p. 77-82.

LEYDESDORFF, Loet. ***The Challenge of Scientometrics****: the development, measurement, and self-organization of scientific communications. 2.ed. Florida, USA: Universal Publishers, 2001.*

LOPES, Maria Margaret. A mesma fé e o mesmo empenho em suas missões científicas e civilizadoras: os museus brasileiros e argentinos do século XIX. **Revista Brasileira de História**. Volume 21. N. 41. 2001.

LUHMANN, Niklas. **Introdução à Teoria de Sistemas**. Tradução de Ana Cristina Arantes Nasser. 2.ed. Petrópolis, RJ: Vozes, 2010.

MAIA, Daniel. **Uma Cartografia de Redes Institucionais: método transdisciplinar e modelagem computacional-imagética da História das Ciências no Brasil**. 2017. Dissertação (Mestrado em História das Ciências, das Técnicas e Epistemologia) - Centro de Ciências Matemáticas e da Natureza, Universidade Federal do Rio de Janeiro. Rio de Janeiro, 2017.

MILIC, Louis T. The Next Step. **Computers and the Humanities**. Volume 1, Issue 1. 1966. p. 3-6. Disponível em: <https://sci-hub.st/10.1007/BF00188010> Acesso em: 15 Mar. 2019.

MORAES FILHO, Alexandre José de Mello. **Revista da Exposição Anthopologica Brazileira**. Rio de Janeiro: Typographia de Pinheiro & C., 1882.

MORETTI, Franco. **Distant Reading**. London, UK: Verso, 2013.

MORTON, Andrew Q., WINSPEAR, Alban D. The Computer and Plato's Seventh Letter. **Computers and the Humanities**. Volume 1, Issue 3. 1967. p. 72-73. Disponível em: <https://sci-hub.st/10.1007/BF00119892>. Acesso em: 22.09.2020.

MUSEU NACIONAL. **O Imperio do Brazil na Exposição Universal de 1876 em Philadelphia**. Rio de Janeiro, Typographia Nacional, 1878.

MUSEU NACIONAL/UFRJ. **Seção de Memória e Arquivo (SEMEAR)**. [S.I.], 2015. Disponível em: <https://www.museunacional.ufrj.br/semear/>. Acesso em: 06 mar. 2021.

NEWTON-SMITH, W.H. ***The Rationality of Science**. New York: Routledge, 2003*.

NICOLESCU, Basarab. **O Manifesto da Transdisciplinaridade**. Tradução de Lucia Pereira de Souza. São Paulo: TRIOM, 2018.

NYHAN, Julianne. PASSAROTTI, Marco (eds.) **One Origin of Digital Humanities**: Fr. Roberto Busa in his Own Words. Switzerland: Springer, 2019.

POPPER, Karl. **Conjectures and Refutations**: the growth of scientific knowledge. New York, USA: Basic Books Publishers, 1962.

PRADO, Hercules A. FERNEDA, Edilson. **Emerging Technologies of Text Mining**: Techniques and Applications. New York, USA: Hershey, 2007.

PRICE, Derek de Solla. **Little Science, Big Science...and Beyond**. New York, USA: Columbia University Press, 1963.

_____. *Network of Scientific Papers.* ***Science****. 1965. N. 149 (3683). p. 510-515.*

PROGHIST LIMITED. **The Programming Historian**. 2021. Página inicial. Disponível em: <https://programminghistorian.org>. Acesso em: 25 jan. 2021.

PUTNAM, George F. Soviets, Quantitative Methods and Digital Computers. **Computers and the Humanities**. Volume 6, N. 1. 1971. p. 23-29. Disponível em: <https://sci-hub.st/10.1007/bf02402320>. Acesso em: 05.04.2020.

RAMSAY, Stephen. **Reading Machines**: Towards an Algorithmic Criticism. Springfield, USA: University of Illinois Press, 2011.

SALTON, G.; WONG, A. On the Role of Words and Phrases in Automatic Text Analysis. **Computers and the Humanities**. Volume 10, Issue 2. 1976. p. 69-87.

SEDELOW, Sally Yeates. The Computer in the Humanities and Fine Arts. **ACM Computing Surveys**. Volume 2, Issue 2. 1970. p. 89-110. Disponível em: <https://sci-hub.st/10.1145/356566.356568>. Acesso em: 05.04.2020.

SILVA, Paulo Vinicius Aprigio da. O Archivos do Museu Nacional e a Institucionalização das Ciências no Século XIX. 2012. (**Dissertação de Mestrado**). Programa de Pós-Graduação em História das Ciências, das Técnicas e Epistemologia da Universidade Federal do Rio de Janeiro, 2012.

SMEDT, Koenraad de. Some Reflections on Studies in Humanities Computing. **Literary and Linguistic Computing**. Volume 17, Issue 1. 2002. p. 89–101.

SMITHSONIAN. **Biodiversity Heritage Library**. Página inicial. [S.l.], 2006. Disponível em: <https://www.biodiversitylibrary.org/>. Acesso em: 20 Jan. 2018.

TERRAS, Melissa; NYHAN, Julianne; VANHOUTTE, Edward (eds.). **Defining Digital Humanities**: a Reader. Farnham, UK: Ashgate, 2013.

THE *GEPHI* CONSORTIUM. ***Gephi***. Página inicial. Versão 0.92. [S.l.], 2010. Disponível em: <https://*Gephi*.org/>. Acesso em: 25 Jan. 2015-

THIBODEAU, Kenneth. Machine Readable Archives and Future History. **Computers and the Humanities**. Volume 10, Issue 2. 1976. p. 89-92.

UFES. **Labic**, 2021. Disponível em: https://www.labic.net/. Acesso em 02 jul. 2021.

UFRJ. **MediaLab.UFRJ**, 2021. Disponível em: https://medialabufrj.net/. Acesso em 25 jun. 2021.

WATTS, Duncan J. **Six Degrees**: the Science of a Connected Age. New York, USA: W. W. Norton, 2003.

_____. **Small Worlds**: the Dynamics of Networks between Order and Randomness. Princeton, USA: Princeton University Press, 1999.

YIN, Robert K. **Pesquisa qualitativa do início ao fim**. Tradução Daniel Bueno. Porto Alegre: Penso, 2016.

Não autenticado Discussão Contribuições Criar u

Página Discussão Imagem

Ler Editar Ver histórico Pesquisar na wiki Wikisource

WIKISOURCE

Página principal
Como participar
Portal comunitário
Mudanças recentes
Obra aleatória
Autor aleatório
Ajuda
Donativos

Ferramentas
Páginas afluentes
Alterações relacionadas
Carregar no Commons
Páginas especiais
Hiperligação permanente
Informações da página
Citar esta página
Recortar imagem

Imprimir/exportar
Criar um livro
Versão para impressão
Descarregar como EPUB
Descarregar como MOBI
Descarregar como PDF
Outros formatos

Línguas

Página:Tratado de paz de Utrecht entre França e Portugal.pdf/9

Esta página foi revisada, mas ainda precisa ser validada

(7)

VIII.

A Fin de prevenir toute occaſion de diſcorde , qui pourroit naiſtre entre les ſujets de la Couronne de France,& ceux de la Couronne de Portugal, Sa Majeſté T.C. deſiſtere pour toujours, comme elle ſe deſiſte dés à préſent par ce Traité dans les termes les plus forts,& les plus authentiques,& avec toutes les clauſes requiſes, comme ſi elles eſtoient inferées icy, tant en ſon nom, q'en celuy de ſe Hoirs, Succeſſeurs & Héritiers de touts droits & prétentions, qu'elle peut & pourra prétendre ſur la proprieté des Terres appellées du *Cap du Nord* & ſituées entre la Riviéres des *Amaſones* & celle de *Iapoc* ou de *Vincent Piſon*, ſans ſe reſerver ou retenir aucune portion deſdites Terres, âfin qu'elles ſoient deſormais poſſedées par Sa Majeſté Portugaiſe, ſes Hoirs,Succeſſeurs & Héritiers avec tous les droits de Souveraiueté,d'abſoluté Puiſſance, & d'inter Domaine,comme faiſant partie de ſes Etats,& qu'elles luy demeurent à perpetuité; ſans que Sa dite Majeſté Portugaiſe, ſes Hoirs, Succeſſeurs & Héritiers puiſſent jamais eſtae troublés dans ladite poſſeſſion par Sa Majeſté T.C. ny par ſe Hoirs,Succeſſeurs,& Héritiers.

IX.

EN conſquence de l'Article précédent Sa Majeſté Portugaiſe pourra faire rebâtir les Forts d'*Araguari*,& de *Camaû*, ou *Maſapa* auſſi bien que tous le autres,qui ont eſté demolis en execution du Traité Proviſionel fait àLisbonne le 4.Mars 1700. entre Sa Majeſté T. C. &

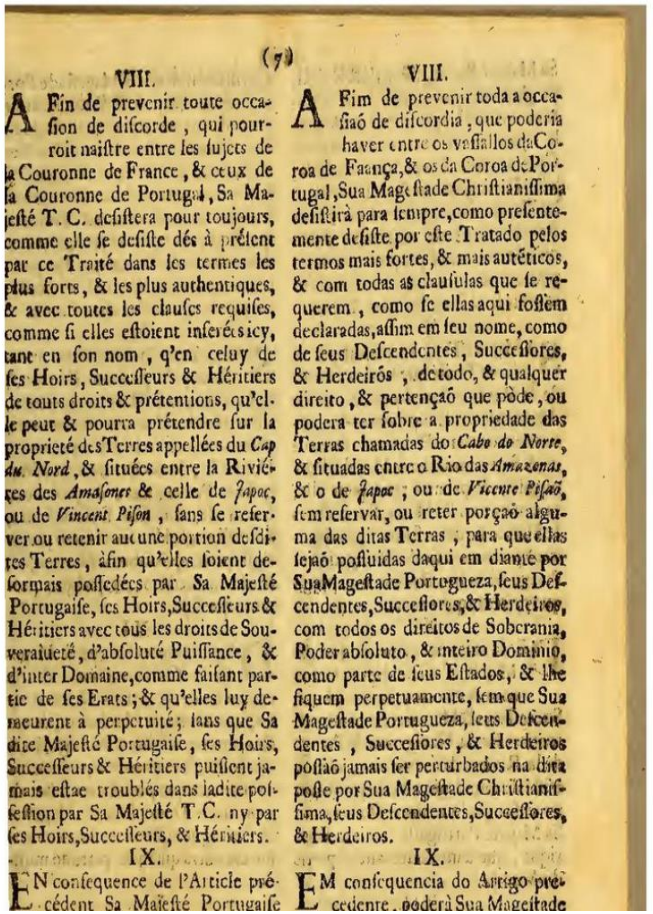
VIII.

A Fim de previnir toda a occaſiaõ de diſcordia,que poderia haver entre os vaſſallos daCoroa de Faança, & os da Coroa de Portugal,Sua Mageſtade Chriſtianiſſima deſiſtirà para ſempre,como preſentemente deſiſte por eſte Tratado pelos termos mais fortes,& mais autẽticos, & com todas as clauſulas que ſe requerem , como ſe ellas aqui foſſem declaradas,aſſim em ſeu nome, como de ſeus Deſcendentes , Succeſſores, & Herdeirõs , de todo, & qualquer direito, & pertençaõ que pòde, ou podera ter ſobre a propriedade das Terras chamadas do *Cabo do Norte*, & ſituadas entre o Rio das *Amazonas*, & o de *Iapoc* , ou de *Vicente Piſaõ*, ſem reſervar, ou reter porçaõ alguma das ditas Terras , para que ellas ſejaõ poſſuidas daqui em diante por SuaMageſtade Portugueza, ſeus Deſcendentes,Succeſſores,& Herdeiros, com todos os direitos de Soberania, Poder abſoluto ,& inteiro Dominio, como parte de ſeus Eſtados, & lhe fiquem perpetuamente, ſem que Sua Mageſtade Portuguesa,ſeus Deſcendentes, Succeſſores ,& Herdeiros poſſaõ jamais ſer perturbados na dita poſſe por Sua Mageſtade Chriſtianiſſima,ſeus Deſcendentes,Succeſſores, & Herdeiros.

IX.

EM conſequencia do Artigo precedente ,poderà Sua Mageſtade Portugueza fazer reedificar os Fortes de *Aguari*, & *Camaù* ,ou *Maſapa*,& os mais que foraõ demolidos em execução do Tratado Proviſional feyto em Lisboa aos 4. de Março de 1700. entre Sua Mageſtade Chriſ-

Sa



(7)

VIII.

A Fin de prevenir toute occaſion de diſcorde , qui pourroit naiſtre entre les ſujets de la Couronne de France , & ceux de la Couronne de Portugal , Sa Majeſté T. C. deſiſtera pour toujours, comme elle ſe deſiſte dés à préſent par ce Traité dans les termes les plus forts , & les plus authentiques, & avec toutes les clauſes requiſes, comme ſi elles eſtoient inferées icy, tant en ſon nom , q'en celuy de ſes Hoirs , Succeſſeurs & Héritiers de touts droits & prétentions, qu'elle peut & pourra prétendre ſur la proprieté des Terres appellées du *Cap du Nord* , & ſituées entre la Riviéres des *Amaſones* & celle de *Iapoc*, ou de *Vincent Piſon* , ſans ſe reſerver ou retenir aucune portion deſdites Terres , âfin qu'elles ſoient deſormais poſſedées par Sa Majeſté Portugaiſe, ſes Hoirs,Succeſſeurs & Héritiers avec tous les droits de Souveraiueté , d'abſoluté Puiſſance , & d'inter Domaine,comme faiſant partie de ſes Erats ; & qu'elles luy demeurent à perpetuité ; ſans que Sa dite Majeſté Portugaiſe, ſes Hoirs, Succeſſeurs & Héritiers puiſſent jamais eſtae troublés dans ladite poſſeſſion par Sa Majeſté T.C. ny par ſes Hoirs,Succeſſeurs, & Héritiers.

IX.

EN conſequence de l'Article précédent Sa Majeſté Portugaiſe

VIII.

A Fim de prevenir toda a occaſiaõ de diſcordia , que poderia haver entre os vaſſallos da Coroa de Faança,& os da Coroa de Portugal ,Sua Mageſtade Chriſtianiſſima deſiſtirà para ſempre,como preſentemente deſiſte por eſte Tratado pelos termos mais fortes, & mais autẽticos, & com todas as clauſulas que ſe requerem , como ſe ellas aqui foſſem declaradas,aſſim em ſeu nome, como de ſeus Deſcendentes , Succeſſores, & Herdeirõs , de todo, & qualquer direito , & pertençaõ que pòde , ou podera ter ſobre a propriedade das Terras chamadas do *Cabo do Norte*, & ſituadas entre o Rio das *Amazonas*, & o de *Iapoc* ; ou de *Vicente Piſaõ*, ſem reſervar, ou reter porçaõ alguma das ditas Terras , para que ellas ſejaõ poſſuidas daqui em diante por SuaMageſtade Portugueza, ſeus Deſcendentes,Succeſſores,& Herdeiros, com todos os direitos de Soberania, Poder abſoluto , & inteiro Dominio, como parte de ſeus Eſtados, & lhe fiquem perpetuamente, ſem que Sua Mageſtade Portugueza, ſeus Deſcendentes , Succeſſores , & Herdeiros poſſaõ jamais ſer perturbados na dita poſſe por Sua Mageſtade Chriſtianiſſima,ſeus Deſcendentes,Succeſſores, & Herdeiros.

IX.

EM conſequencia do Artigo precedente , poderà Sua Mageſtade

ANEXO 1

ANEXO 2

Reprodução do site. https://humanidadesdigitais.org/manifesto-das-humanidades-digitais/

Manifesto das Humanidades Digitais

Poster em Português. Tradução de Hervé Théry. Poster confeccionado pela equipe hypotheses.org. Gentilmente enviado para o HD.br por Marie Pellen

Publicado originalmente no THATCamp 2012

Veja também:
| Um breve panorama
| Pequeno roteiro de leituras

Manifesto das Humanidades Digitais

26 de março de 2011

Por Marin Dacos

Contexto

Nós, atores ou observadores das digital humanities *(humanas digitais), reunimo-nos em Paris no momento de realização do THATCamp dos dias 18 e 19 de maio de 2010. Durante estes dois dias, discutimos, trocamos ideias, refletimos juntos sobre as* digital humanities *e tentamos imaginar e inventar o que poderiam ser.*

Poster em Inglês. Texto e poster original do THATcamp 2012.
Outras versões em http://tcp.hypotheses.org

Ao final desses dois dias, que consideramos apenas uma etapa, propusemos um manifesto das digital humanities *às comunidades de pesquisa e a quem participa da criação, da edição, da valorização ou da conservação dos conhecimentos.*

I. Definição

1. A opção da sociedade pelo digital altera e questiona as condições de produção e divulgação dos conhecimentos.

2. Para nós, as *digital humanities* referem-se ao conjunto das Ciências humanas e sociais, às Artes e às Letras. As humanas digitais não negam o passado, apoiam-se, pelo contrário, no conjunto dos paradigmas, *savoir-faire* e conhecimentos próprios dessas disciplinas, mobilizando simultaneamente os instrumentos e as perspetivas singulares do mundo digital.

3. As *digital humanities* designam uma transdisciplina, portadora dos métodos, dos dispositivos e das perspetivas heurísticas ligadas ao digital no domínio das Ciências humanas e sociais.

II. Situação

4. Constatamos:

– que se multiplicaram experimentações no domínio do digital em Ciências humanas e sociais nos últimos cinquenta anos; que emergiram, mais recentemente, centros das *digital humanities*, que são todos, atualmente, protótipos ou lugares de aplicação específica de uma abordagem das *digital humanities*;

– que o digital induz uma presença mais forte dos aspectos técnicos e econômicos na pesquisa; que esta obrigação é uma oportunidade para fazer evoluir o trabalho coletivo;

– que existem diversos métodos provados, conhecidos e compartilhados desigualmente;

– que existem múltiplas comunidades específicas, oriundas de interesses por diversas práticas, instrumentos ou objetos transversais (codificação de fontes textuais; sistemas de informação geográfica; lexicometria; digitalização do patrimônio cultural, científico e técnico; cartografia da web; garimpagem de dados; 3D; arquivos orais; artes e literaturas digitais e hipermediáticas; etc.), que convergem atualmente para formar o campo das *digital humanities*.

III. Declaração

5. Nós, atores do *digital humanities*, constituímo-nos numa comunidade de prática solidária, aberta, acolhedora e de livre acesso.

6. Somos uma comunidade sem fronteiras. Somos uma comunidade multilingue e multidisciplinar.

7. Além da esfera acadêmica, nossos objetivos são o progresso do conhecimento, o reforço da qualidade da pesquisa em nossas disciplinas, e o enriquecimento do saber e do patrimônio coletivo.

8. Convocamos à integração da cultura digital na definição da cultura geral do século XXI.

IV. Orientações

9. Nós lançamos um apelo ao acesso livre aos dados e os metadados. Estes devem ser documentados e interoperáveis, tanto técnica como conceitualmente.

10. Somos a favor da divulgação, da circulação e do livre enriquecimento dos métodos, do código, dos formatos e dos resultados da pesquisa.

11. Convocamos à integração de formações em *digital humanities* nos curriculos em Ciências humanos e sociais, Artes e Letras. Desejamos igualmente a criação de diplomas em *digital humanities* e o desenvolvimento de formações profissionais específicas. Por último, desejamos que estas competências sejam consideradas nos recrutamentos e nas evoluções de carreira.

12. Comprometemo-nos com a edificação de uma competência coletiva que se apoie em um vocabulário comum, competência coletiva que procede do trabalho do conjunto dos atores. Essa competência coletiva deve tornar-se um bem comum. Constitui uma oportunidade científica, mas também uma oportunidade de inserção profissional, em todos os setores.

13. Desejamos participar da definição e da divulgação de boas práticas, correspondentes a necessidades disciplinares e transdisciplinares identificadas, que são evolutivas e procedentes de um debate e de um consenso nas comunidades interessadas. A abertura fundamental das humanas digitais assegura no entanto uma abordagem pragmática dos protocolos e das visões, que mantém o direito à coexistência de métodos diferentes e concorrentes, para o enriquecimento da reflexão e das práticas.

14. Chamamos à construção de ciber-infra-estruturas evolutivas que respondam a necessidades reais. Estas ciber-infra-estruturas construir-se-ão de maneiras iterativas, apoiando-se sobre a constatação de métodos e de abordagens comprovadas nas comunidades de pesquisa.

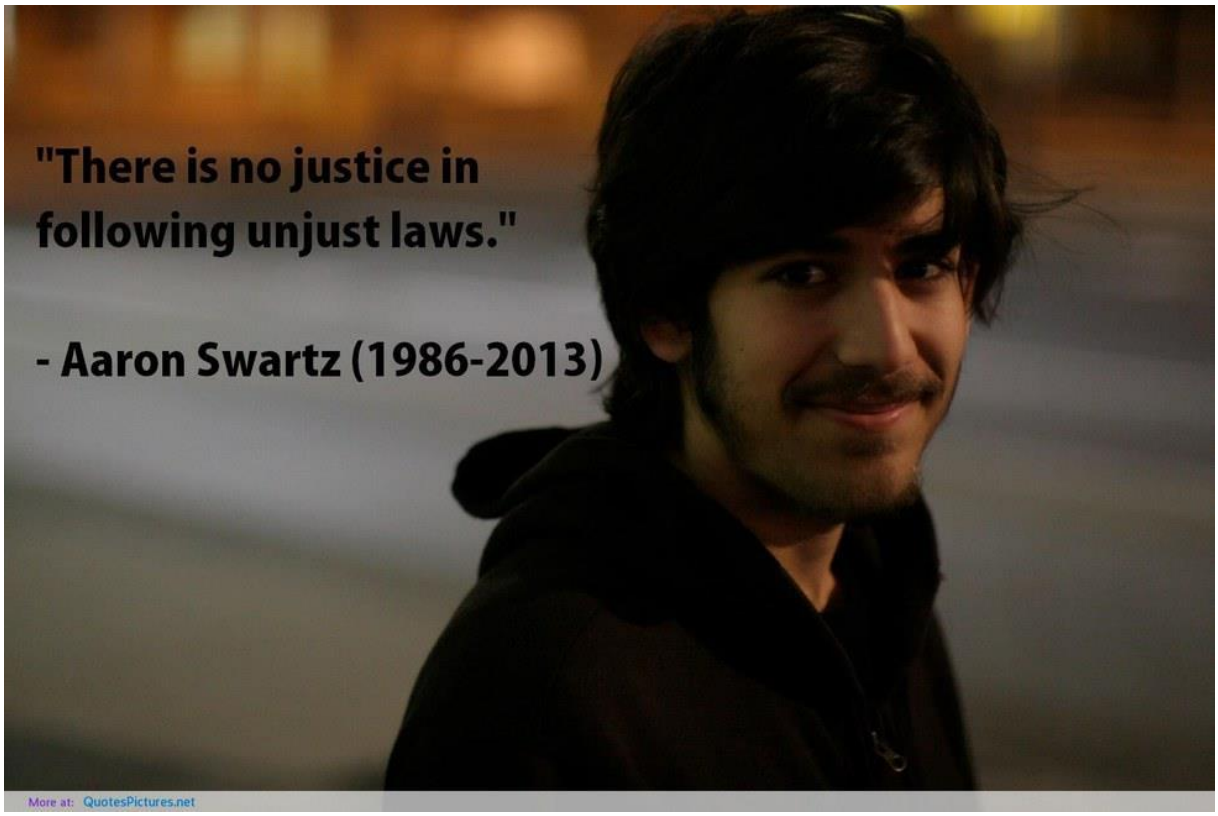
Juntem-se a nós!

Traduction par Hervé Théry

ANEXO 3

GUERILLA OPEN ACCESS MANIFESTO

(MANIFESTO DA GUERRILHA DO LIVRE ACESSO).
Tradução e Adaptação por Daniel Maia



Não há justiça ao seguir leis injustas. Aaron Swartz (1986-2013).

Informação é poder. Mas, como todo poder, há quem queira mantê-lo para si. Todo o patrimônio científico e cultural mundial, publicado ao longo dos séculos em livros e periódicos, está cada vez mais sendo digitalizado e bloqueado por um punhado de empresas privadas. Quer ler os artigos que apresentam os resultados mais famosos das ciências? Você precisará enviar quantias enormes para editores como Reed Elsevier.

Existem aqueles que estão lutando para mudar isso. O Movimento Acesso Aberto tem lutado bravamente para garantir que os cientistas não assinem seus direitos autorais, mas, em vez disso, certifique-se de que seu trabalho seja publicado na Internet, em

termos que permitem o acesso de qualquer pessoa. Mas mesmo nos melhores cenários, seu trabalho só se aplicará a coisas publicadas no futuro. Tudo até agora terá sido perdido.

É um preço muito alto a pagar. Forçando acadêmicos a pagarem para ler o trabalho de seus colegas? Verificando bibliotecas inteiras, mas permitindo apenas que o pessoal do Google as leia? Fornecendo artigos científicos para aqueles em universidades de elite no Primeiro Mundo, mas não para crianças no Sul Global? É ultrajante e inaceitável.

“Eu concordo”, muitos dizem, “mas o que podemos fazer? As empresas detêm os direitos autorais, elas ganham enormes quantias de dinheiro cobrando pelo acesso, e isso é perfeitamente legal - não há nada que possamos fazer para impedi-las. ” Mas há algo que podemos, algo que já está sendo feito: podemos contra-atacar.

Aqueles com acesso a esses recursos - estudantes, bibliotecários, cientistas - receberam um privilégio. Você pode se alimentar deste banquete de conhecimento enquanto o resto do mundo está trancado do lado de fora. Mas você não precisa - na verdade, moralmente, você não pode - manter esse privilégio para si mesmo. Você tem o dever de compartilhá-lo com o mundo. E você tem: trocar senhas com colegas, preencher solicitações de download de amigos.

Enquanto isso, aqueles que foram bloqueados não ficam de braços cruzados. Você tem se esgueirado por buracos e pular cercas, liberando as informações trancadas pelos editores e compartilhando-as com seus amigos.

Mas toda essa ação acontece no escuro, escondido no subsolo. É chamado de roubo ou pirataria, como se compartilhar uma riqueza de conhecimento fosse o equivalente moral a saquear um navio e assassinar sua tripulação. Mas compartilhar não é imoral - é um imperativo moral. Apenas aqueles cegos pela ganância se recusariam a deixar um amigo fazer uma cópia.

É claro que as grandes corporações estão cegas pela ganância. As leis sob as quais operam exigem isso - seus acionistas se revoltariam com qualquer coisa menos. E os políticos que eles compraram de volta, aprovando leis que lhes dão o poder exclusivo de decidir quem pode fazer cópias.

Não há justiça em seguir leis injustas. É hora de vir à luz e, na grande tradição da desobediência civil, declarar nossa oposição a esse roubo privado da cultura pública.

Precisamos levar as informações, onde quer que estejam armazenadas, fazer nossas cópias e compartilhá-las com o mundo. Precisamos pegar o material que não está protegido por direitos autorais e adicioná-lo ao arquivo. Precisamos comprar bancos de dados secretos e colocá-los na web. Precisamos fazer download científico diários

e carregue-os em redes de compartilhamento de arquivos. Precisamos lutar pelo *Guerilla Open Access*.

Com o suficiente de nós, em todo o mundo, não enviaremos apenas uma mensagem forte de oposição à privatização do conhecimento - faremos disso uma coisa do passado. Você vai se juntar a nós?

Aaron Swartz

Julho de 2008, Eremo, Itália

Original em:

https://ia600605.us.archive.org/15/items/GuerillaOpenAccessManifesto/Goamjuly2008.pdf

ANEXO 4

CARTA DA TRANSDISCIPLINARIDADE

Preâmbulo

Considerando que a proliferação atual das disciplinas acadêmicas e não acadêmicas conduz a um crescimento exponencial do saber, o que torna impossível uma visão global do ser humano;

Considerando que somente uma inteligência que leve em consideração a dimensão planetária dos conflitos atuais poderá enfrentar a complexidade do nosso mundo e o desafio contemporâneo de autodestruição material e espiritual de nossa espécie;

Considerando que a vida está fortemente ameaçada por uma tecnociência triunfante, que só obedece à lógica assustadora da eficácia pela eficácia;

Considerando que a ruptura contemporânea entre um saber cada vez mais acumulativo e um ser interior cada vez mais empobrecido leva à uma ascensão de um novo obscurantismo cujas consequências, no plano individual e social, são incalculáveis;

Considerando que o crescimento dos saberes, sem precedente na história, aumenta a desigualdade entre os aqueles que os possuem e os que deles estão desprovidos, gerando assim desigualdades crescentes no seio dos povos e entre as nações do nosso planeta;

Considerando, ao mesmo tempo, que todos os desafios enunciados tem sua contrapartida de esperança e que o crescimento extraordinário dos saberes pode conduzir, a longo prazo, a uma mutação comparável à passagem dos hominídeos à espécie humana;

Considerando o que precede, os participantes do Primeiro Congresso Mundial de Transdisciplinaridade (Convento da Arábia, Portugal, 2 a 7 de novembro de 1994) adoram a presente Carta, que contém um conjunto de princípios fundamentais da comunidade dos espíritos transdisciplinares, constituindo um contrato moral que todo signatário dessa Carta faz consigo mesmo, livre de qualquer espécie de pressão jurídica ou institucional.

Artigo 1

Toda e qualquer tentativa de reduzir o ser humano a uma definição e de dissolvê-lo em estruturas formais, sejam quais forem, é incompatível com a visão transdisciplinar.

Artigo 2

O reconhecimento da existência de diferentes Níveis de Realidade, regido por lógicas diferentes, é inerente à atitude transdisciplinar. Toda tentativa de reduzir a Realidade a um só nível, regido por uma lógica única, são se situa no campo da transdisciplinaridade.

Artigo 3

A Transdisciplinaridade é complementar à abordagem disciplinar; ela faz emergir novos dados a partir da confrontação das disciplinas que os articulam entre si; ela nos oferece uma nova visão da Natureza e da Realidade. A transdisciplinaridade não procura o domínio de várias disciplinas, mas a abertura de todas as disciplinas ao que as une e as ultrapassa.

Artigo 4

A pedra angular da transdisciplinaridade reside na unificação semântica e operativa das acepções através e além das disciplinas. Ela pressupõe uma racionalidade aberta, mediante um novo olhar sobre a relatividade das noções de definição e de objetividade'. O formalismo excessivo, a rigidez das definições e o exagero da objetividade, incluindo-se a exclusão do sujeito, conduzem ao empobrecimento.

Artigo 5

A visão transdisciplinar é resolutamente aberta na medida em que ela ultrapassa o campo das ciências exatas devido ao seu diálogo e sua reconciliação, não somente com as ciências humanas, mas também com a arte, a literatura, a poesia e a experiência interior.

Artigo 6

Com relação à interdisciplinaridade e à multidisciplinaridade, a transdisciplinaridade é multirreferencial e multidimensional. Embora levando em conta os conceitos de tempo e de História, a transdisciplinaridade não exclui a existência de um horizonte transhistórico.

Artigo 7

A transdisciplinaridade não constitui nem uma nova religião, nem uma nova filosofia, nem uma nova metafísica, nem uma ciência da ciência.

Artigo 8

A dignidade do ser humano também é de ordem cósmica e planetária. O aparecimento do ser humano na Terra é uma das etapas da história do Universo. O reconhecimento da Terra como pátria é um dos imperativos da transdisciplinaridade. Todo ser humano

tem direito a uma nacionalidade, mas a título de Habitante da Terra, ele é ao mesmo tempo um ser transnacional. O reconhecimento pelo direito internacional da dupla cidadania pertencer a uma nação e à Terra – constitui um dos objetivos da pesquisa transdisciplinar.

Artigo 9

A transdisciplinaridade conduz a uma atitude aberta em relação aos mitos, religiões e temas afins, que os respeitam em um espírito transdisciplinar.

Artigo 10

Não existe um lugar cultural privilegiado de onde se possa julgar as outras culturas. A abordagem transdisciplinar é, ela própria, transcultural.

Artigo 11

Uma educação autêntica não pode privilegiar abstração no conhecimento. Ela deve ensinar a contextualizar, concretizar e globalizar. A educação transdisciplinar reavalia o papel da intuição, do imaginário, da sensibilidade e do corpo na transmissão do conhecimento.

Artigo 12

A elaboração de uma economia transdisciplinar está baseada no postulado de que a economia deve estar a serviço do ser humano e não o inverso.

Artigo 13

A ética transdisciplinar recusa toda e qualquer atitude que se negue ao diálogo e à discussão, qualquer que seja a sua origem – de ordem ideológica, cientificista, religiosa, econômica, política, filosófica. O saber compartilhado deve levar a uma compreensão compartilhada, baseado no respeito absoluto às alteridades unidas pela vida comum numa só e mesma Terra.

Artigo 14

Rigor, abertura e tolerância são as características fundamentais da atitude e da visão transdisciplinares. O rigor na argumentação que leva em conta todos os dados é a melhor barreira em relação aos possíveis desvios. A abertura comporta a aceitação do desconhecido, do inesperado e do imprevisível. A tolerância é o reconhecimento do direito às ideias e verdades contrárias às nossas.

Artigo final

A presente Carta da Transdisciplinaridade foi adotada pelos participantes do Primeiro Congresso Mundial de Transdisciplinaridade e não reivindica nenhuma outra autoridade além de sua obra e da sua atividade.

Segundo os procedimentos que serão definidos em acordo com as mentes transdisciplinares de todos os países, esta Carta está aberta à assinatura de qualquer ser humano interessado em promover nacional, internacional e transnacionalmente as medidas progressivas para a aplicação destes artigos na vida cotidiana.

Convento da Arrábida, 6 de novembro de 1994.

Comitê de Redação
Lima de Freitas, Edgar Morin e Basarab Nicolescu.