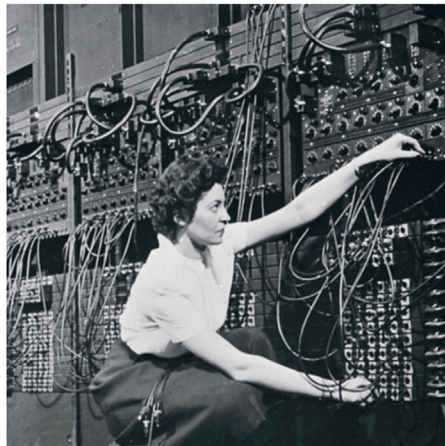


PRÁTICAS DA HISTÓRIA

JOURNAL ON THEORY, HISTORIOGRAPHY,
AND USES OF THE PAST

Nº 14 - 2022



História digital, mineração de texto e análise textual: algumas possibilidades através da utilização de ferramentas oferecidas pelo Perseus Project

Ismael Wolf

Práticas da História, n.º 14 (2022): 151-168

www.praticasdahistoria.pt

Ismael Wolf

História digital, mineração de texto e análise textual: algumas possibilidades através da utilização de ferramentas oferecidas pelo Perseus Project

O presente ensaio trata inicialmente de uma breve explanação sobre o que são as Humanidades Digitais, se utilizando para isso de bibliografia recente sobre essa temática. Após essa explanação inicial, é realizado um aprofundamento sobre o conceito de mineração de texto e é também apresentado um breve relato/comentário crítico tratando da utilização do Perseus enquanto ferramenta inicial para mineração de texto. São apresentadas possibilidades e limitações constatadas na utilização deste ferramental.

Palavras-chave: Humanidades digitais, Mineração de texto, Perseus Project, História digital.

Digital history, text mining and text analysis: some possibilities through the use of Perseus Project tools

This essay is initially a brief explanation of what the Digital Humanities are, using recent bibliography on this subject. After this initial explanation, an in-depth look at the concept of text mining is carried out and a brief report/critical comment dealing with the use of Perseus as an initial tool for text mining is also presented. Possibilities and limitations found in the use of this tool are presented.

Keywords: Digital humanities, Text mining, Perseus Project, Digital history.

História digital, mineração de texto e análise textual: algumas possibilidades através da utilização de ferramentas oferecidas pelo Perseus Project

Ismael Wolf*

Humanidades Digitais (*Digital Humanities*)

Nos últimos anos temos observado um crescente avanço do que chamamos de Humanidade Digitais. Obviamente que este avanço tem sido não apenas impulsionado pelo aumento de interesse pelas Ciências Humanas, mas sim pelos grandes saltos tecnológicos que temos vivido. Jamais a humanidade experimentou tantos avanços e descobertas tecnológicas. Desta forma é natural que essas novas tecnologias sejam utilizadas para os mais diversos fins. Devemos lembrar que as Humanidades Digitais têm um caráter interdisciplinar atingindo diversas disciplinas e contribuindo para a troca entre elas. Alguns autores preferem utilizar, inclusive, o termo transdisciplinar. Segundo Rosenfield:

A transdisciplinaridade radicaliza as normas e práticas disciplinares existentes e permite aos pesquisadores ir além de suas disciplinas originárias, usando uma estrutura conceitual compartilhada que reúne conceitos, teorias e abordagens de várias disciplinas em algo novo que os transcende a todos¹. (Tradução nossa.)

* Ismael Wolf (wolf_ismael@yahoo.co.uk; ismael.wolf@edu.unirio.br). Doutorando no Programa de Pós-Graduação em História da Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Av. Pasteur, 458 - Urca - Rio de Janeiro – RJ, Brasil - CEP 22290-240. Professor titular de História na Rede Municipal de Educação de Gravataí – RS, Brasil. Art. Ensaio elaborado durante a disciplina de História e Arqueologia Digitais, ministrada pela Profa. Dra. Adriene Baron Tacla, no Programa de Pós-Graduação em História da Universidade Federal Fluminense. Texto original: 10-12-2021; Versão revista: 14-6-2022; Aceite para publicação: 22-6-2022.

1 Patricia L. Rosenfield, “The Potential of Transdisciplinary Research for Sustaining and Ex-

Esta nova perspectiva transdisciplinar cria uma atmosfera de trabalho que incentiva e traz a necessidade de se criarem redes e grupos de trabalho. Em geral, os projetos de Humanidades Digitais demandam o trabalho colaborativo de todo um time de pesquisadores. Mas, afinal, as Humanidades Digitais substituem as pesquisas tradicionais? De acordo com Evans:

A pesquisa nas humanidades digitais é movida pelo mesmo desejo de saber que as humanidades tradicionais, mas através do novo “saber” que a presença da computação deu, os tópicos e *áreas de pesquisa, os métodos e os resultados mudaram, mas as humanidades são as humanidades, sejam digitais ou tradicionais. Esta é a segunda onda* – as atividades e funções das humanidades sendo atendidas por métodos digitais, bem como por meio da leitura tradicional, mas levando a uma mudança nos métodos e na produção das humanidades². (Tradução nossa.)

Percebemos com isso que não há uma substituição, já que as Humanidades continuam sendo as Humanidades. O que ocorre é que a tecnologia é colocada a serviço das Ciências Humanas, possibilitando a utilização de novos métodos e a possibilidade de novos tipos de questionamentos e hipóteses.

tending Linkages between the Health and Social Sciences”, *Social Science & Medicine* 35, n.º 11 (1992): 1351, citado em Yu-wei Lin, “Transdisciplinarity and Digital Humanities: Lesson Learned from Developing Text-Mining Tools for Textual Analysis”, in *Understanding Digital Humanities*, org. David M. Berry (Basingstoke: Palgrave Macmillan, 2012), 298. “Transdisciplinarity radicalises existing disciplinary norms and practices and allows researchers to go beyond their parent disciplines, using a shared conceptual framework that draws together concepts, theories, and approaches from various disciplines into something new that transcends them all.”

2 Leighton Evans e Sian Rees, “An Interpretation of Digital Humanities”, in *Understanding Digital Humanities*, org. David Berry (Basingstoke: Palgrave Macmillan, 2012), 31. “Research in the digital humanities is driven by the same desire to know as traditional humanities, but that through the new ‘knowing-how’ that the presence of computation has given, the topics and areas of research, the methods and the results have changed, but the humanities are the humanities, whether digital or traditional. This is the second wave – the activities and functions of the humanities being served through digital methods, as well as through traditional close reading, but leading to a change in the methods and output of the humanities.”

Inicialmente a utilização do computador neste tipo de pesquisa, por exemplo, se dava apenas para a realização de análises quantitativas. Pesquisadores de história econômica se utilizaram dessas novas ferramentas para elaborar e comprovar suas hipóteses³. Com o tempo, as possibilidades se ampliaram e, atualmente, já é possível utilizar certas ferramentas para se fazer também análises qualitativas.

Com o desenvolvimento das Humanidades Digitais se percebeu também a necessidade de criar um “movimento de dados abertos”⁴. O compartilhamento de informações através de grandes bancos de dados tornou-se uma realidade. Claro que ainda há muito o que se melhorar e expandir neste sentido, mas pode-se perceber que este movimento tem crescido. Alguns países, inclusive, têm dado preferência ao financiamento de projetos que tenham uma base de dados aberta. A dificuldade muitas vezes tem se dado em conseguir os recursos para manter esses projetos, já que na maioria das vezes o investimento financeiro para compra, desenvolvimento e manutenção é extremamente caro. Mesmo com as dificuldades relatadas, projetos interessantes de arquivamento de dados têm surgido. Podemos citar, por exemplo, o arquivamento de dados da Web que tem sido realizado pelo Projeto Internet Archive⁵, onde diversas páginas da Web têm sido salvas através de *backups* contínuos. No entanto, algumas dificuldades são percebidas como, por exemplo, *hyperlinks* de páginas desativadas e o excesso de informação a ser armazenada. Uma cópia fiel da página nunca estará disponível, apenas um registro de um determinado momento. Projetos como o Internet Archive possibilitam que informações importantes da Web não sejam perdidas, ou pelo menos parte delas. A Web enquanto plataforma⁶ tem nos proporcionado uma oferta crescente e diversa de conteúdo. Esse excesso de informação também acaba sendo um problema para que

3 Roger Middleton e Peter Wardley, “Information Technology in Economic and Social History: The Computer as Philosopher’s Stone or Pandora’s Box?”, *The Economic History Review* 43, n.º 4 (1990): 667-96.

4 Laia Pujol Priego, Jonathan Wareham e Angelo Kenneth S. Romasanta, “The Puzzle of Sharing Scientific Data”. *Industry and Innovation*, 29, n.º 2 (2022): 219-50.

5 Internet Archive, acesso em 30 nov. 2021, <https://archive.org/>.

6 Ver *Web 2.0*.

ocorra o arquivamento de certos dados para pesquisas futuras, já que escolhas precisam de ser realizadas e com isso muitos dados acabam se perdendo, o que se torna problemático para a conservação do patrimônio digital.

Ferramentas importantes da Web, como o Google Acadêmico e a Academia.edu, têm contribuído para o compartilhamento de informações, dados, artigos. Bancos de dados como o Celtic Coindex⁷, por exemplo, também têm ajudado neste sentido de compartilhamento. Bancos de dados, buscadores e ferramentas da Web acabam contribuindo para o desenvolvimento de áreas como a Arqueologia e a História. Buscas por dados arqueológicos podem ser realizadas com certa facilidade pela Web, mas isso não significa que tudo funcione perfeitamente. Algoritmos como os utilizados pelo Google estabelecem padrões individuais de acordo com o perfil de cada usuário, o que pode fazer com que a pesquisa omita informações que poderiam ser importantes para o pesquisador. Há ainda o problema de direcionamento de acordo com os interesses das grandes empresas da Web. Pesquisas sobre Arqueologia podem ser impactadas pelos interesses do turismo, por exemplo, onde talvez o algoritmo direcione a pesquisa para sítios mais famosos e patrocinados, patrimônios da humanidade, etc. Outras ferramentas da Web como o Google Maps, o Google Earth e as mídias sociais podem fornecer dados importantes de acordo com o tipo de pesquisa realizada.

A Arqueologia tem se desenvolvido fortemente entre as Humanidades Digitais. Embora a Arqueologia sempre tenha tido uma interação com as ciências duras, ela tem sido impactada pelas novas tecnologias. Através da Arqueologia Digital o arqueólogo pode recorrer a novos e diferentes métodos antes inacessíveis. Entre as perspectivas da Arqueologia Digital, podemos citar a manutenção da importância do trabalho de campo, que permanece, mas que também pode ser realizado em parte remotamente, as análises quantitativas e qualitativas, métodos de GIS⁸, ensino e divulgação. A Arqueologia Digital não é um novo

⁷ Banco de dados de moedas celtas.

⁸ Geographic Information System.

paradigma ou uma nova ciência. Apenas apresenta novas ferramentas e novas maneiras de lidar com os objetos, muitas vezes antigos e já analisados anteriormente com o uso de outras ferramentas. Vários bancos de dados também têm contribuído para o compartilhamento de informações. Podemos citar alguns importantes como Archaeology Data Service (ADS)⁹, History Data Service (HDS)¹⁰, The Digital Archaeological Record (TDAR)¹¹, Pelagios.org Commons¹², Digital Augustan Rome¹³ e The Portable Antiquities Scheme¹⁴, cada um destes com suas especificidades.

Dentre as novas possibilidades trazidas pela Arqueologia Digital estão métodos de análise de redes (SNA), em que podem ser percebidos pontos de contato, interação e distanciamento, através de análises estatísticas, biológicas, genéticas, linguísticas e geográficas, por exemplo¹⁵. A reconstituição de comunidades através de visualização 3D¹⁶ e ferramentas como impressoras 3D também têm trazido grande contribuição para pesquisas e para o ensino. Há também outras ferramentas importantes para análise de mapas. Podemos citar o exemplo do Pelagios como uma ferramenta muito interessante para a análise do mundo romano.

Outra preocupação das Humanidades Digitais se dá através do processo de digitalização de documentos. Diferentemente dos nascidos digitais, os documentos digitalizados necessitam de certos cuidados especiais. É importante frisar que é necessário que os pesquisadores que lidam com esses documentos compreendam como ocorreram os proces-

9 Archaeology Data Service, acesso em 30 nov. 2021, <http://archaeologydataservice.ac.uk/>.

10 UK Data Archive, acesso em 30 nov. 2021, <http://www.data-archive.ac.uk/about/projects/hds/>.

11 The Digital Archaeological Record, acesso em 30 nov. 2021, <http://core.tdar.org/>.

12 Pelagios, acesso em 30 nov. 2021, <https://pelagios.org/>.

13 Digital Augustan Roman, acesso em 30 nov. 2021, <http://www.digitalaugustanrome.org/>.

14 Portable Antiquities Scheme, acesso em 30 nov. 2021, <https://finds.org.uk/>.

15 John Edward Terrel, "Social Network Analysis and the Practice of History", in *Network Analysis in Archaeology: New Approaches to Regional Interaction*, org. Carl Knappett (Oxford: Oxford University Press, 2013), 17-42.

16 Bettina Arnold, Kevin Garstki e Matthew L. Murray, "Reconstituting Community: 3D Visualization and Early Iron Age Social Organization in the Heuneburg Mortuary Landscape", *Journal of Archaeological Science*, 54 (2015): 23-30.

sos de digitalização. Quem digitalizou? Esse era um *expert* da área? Conhecia a língua original do documento? Em que época ocorreu a digitalização? Como foi digitalizado? Todas estas perguntas são importantes para o pesquisador que trabalha com documentos digitalizados. E aquele que digitaliza um livro antigo, por exemplo, precisa de estar atento a todos os detalhes. Informações que antes eram irrelevantes como a capa, a contracapa, as margens das folhas, o material de que é feito o livro, hoje são de extrema importância para certos tipos de pesquisa. Por este motivo é importante que não apenas os textos sejam preservados, mas todos os componentes de um livro ou pergaminho¹⁷.

Além de livros e pergaminhos digitalizados percebemos também uma crescente de propostas de livros já nascidos em formato digital. Muitos desses são livros eletrônicos de história. Esse tipo de proposta possibilita diversas situações que não são possíveis com os livros normais. Com os livros eletrônicos é possível ter acesso a fontes, arquivos de áudio e vídeos, por exemplo. Fora o fato de que a capacidade de armazenamento de informação é muito superior. O uso de elementos multimídia, a possibilidade de mover-se rapidamente de um corpo de informação para outro através de *hiperlinks* e a possibilidade de navegar rapidamente dentro do livro através de uma busca por palavras são algumas das vantagens deste formato, que não necessariamente sentenciam ao esquecimento o modelo de livro tradicional, mas nos fornece mais uma possibilidade dentre o leque de novas possibilidades oferecidas pelas Humanidades Digitais¹⁸.

Após ter apresentado uma breve explanação sobre as Humanidades Digitais e também algumas possibilidades trazidas pelas mesmas, abordarei no próximo tópico a utilização de ferramentas de mineração de texto, assim como sua importância para a pesquisa em Humanidades Digitais.

17 Harold Thwaites, "Digital Heritage: What Happens When We Digitize Everything?", in *Visual Heritage in the Digital Age*, org. Henry Chapman, Eugene Ch'ng e Vincent Gaffney (Londres: Springer-Verlag, 2013), 327-48.

18 Steve Brier, "Historians and Hypertext: Is It More Than Hype?", in *Clio Wired: The Future of The Past in The Digital Age*, org. Roy Rosenzweig (Nova Iorque: Columbia University Press, 2011), 85-91.

Mineração de texto (*Text Mining*)

Um dos tipos de ferramentas e técnicas que têm sido utilizadas nas Humanidades Digitais é a chamada mineração de texto¹⁹. Embora esteja sendo usada pela área de Humanidades, a mineração de texto inicialmente surgiu como uma ferramenta para auxiliar pesquisadores de outras áreas. Mas afinal o que esse termo significa? Alguns autores podem nos oferecer a resposta para esta pergunta. De acordo com Ambrósio e Morais:

Mineração de textos (*Text Mining*) é um processo de descoberta de conhecimento, que utiliza técnicas de análise e extração de dados a partir de textos, frases ou apenas palavras. Envolve a aplicação de algoritmos computacionais que processam textos e identificam informações úteis e implícitas [...] sendo que suas principais contribuições estão relacionadas à busca de informações específicas em documentos, à análise qualitativa e quantitativa de grandes volumes de textos, e a melhor compreensão do conteúdo disponível em documentos textuais²⁰.

Todavia, diferentemente de uma simples pesquisa como as que fazemos em buscadores como o Google, por exemplo, a mineração de texto tem como objetivo encontrar informações até então desconhecidas. Não se está buscando algo já pronto e selado, mas dados que podem levar a formular algo novo. Sobre isso nos fala Hearst:

A mineração de texto é diferente do que conhecemos na pesquisa na *web*. Na pesquisa, o usuário normalmente está procurando por algo que já é conhecido e foi escrito por

19 Em inglês: *Text-mining*.

20 Ana Paula L. Ambrósio e Edilson Andrade Martins Morais, “Mineração de textos”, Instituto de Informática da UFG, acesso em 30 nov. 2021, http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf.

outra pessoa. O problema é deixar de lado todo o material que, atualmente, *não é relevante para as suas necessidades para encontrar as informações relevantes. Na mineração de texto, o objetivo é descobrir informações até então desconhecidas, algo que ninguém ainda sabe e, portanto, ainda não poderia ter escrito*²¹. (Tradução nossa.)

Aliado a isso temos um processo que une profissionais de diferentes disciplinas. Não apenas o profissional que irá fazer a pesquisa, mas é necessário que se trabalhe em conjunto com profissionais de TI²², por exemplo. Apenas este trabalho conjunto irá trazer resultados plenamente satisfatórios, já que o profissional de TI poderá contribuir com seu conhecimento para criar um algoritmo que permita que o pesquisador das humanidades possa chegar aos resultados que espera encontrar. Segundo a página da UK JISC-funded National Centre for Text Mining (NaCTeM):

A mineração de texto envolve a aplicação de técnicas de áreas como recuperação de informação, processamento de linguagem natural, extração de informação e mineração de dados. Esses vários estágios de um processo de mineração de texto podem ser combinados em um único fluxo de trabalho²³. (Tradução nossa.)

21 Marti Hearst, “What Is Text Mining?”, SIMS, UC Berkeley, acesso em 30 nov. 2021, <http://people.ischool.berkeley.edu/~hearst/text-mining.html>. “Text mining is different from what we’re familiar with in web search. In search, the user is typically looking for something that is already known and has been written by someone else. The problem is pushing aside all the material that currently isn’t relevant to your needs in order to find the relevant information. In text mining, the goal is to discover heretofore unknown information, something that no one yet knows and so could not have yet written down.”

22 Tecnologia da Informação.

23 UK JISC-funded National Centre for Text Mining (NaCTeM), acesso em 16 jan. 2017, <http://www.jisc.ac.uk/publications/briefingpapers/2008/bptextminingv2.aspx>. “Text mining involves the application of techniques from areas such as information retrieval, natural language processing, information extraction and data mining. These various stages of a text-mining process can be combined into a single workflow.”

É importante lembrar que não existe apenas uma forma de mineração de textos. Diferentes ferramentas podem proporcionar diferentes tipos de análises textuais e diversos tipos de levantamento de dados. Esse é um dos motivos do termo mineração de texto ter sido considerado como um termo “guarda-chuva” por Lin:

Diante disso, a mineração de texto pode ser entendida como um termo abrangente para incorporar e implementar uma ampla gama de ferramentas ou técnicas (algoritmos, métodos), incluindo mineração de dados, aprendizado de máquina, processamento de linguagem natural, inteligência artificial, agrupamento, mineração de conhecimento e análise de texto, linguística computacional, análise de conteúdo e análise de sentimento e assim por diante, em um grande corpo de textos (geralmente uma enorme coleção de documentos) para apoiar a tomada de decisões dos usuários. Assim como peças de Lego, existe um conjunto de componentes em campo que podem ser montados e reconfigurados para fins de tarefas dos usuários do domínio²⁴. (Tradução nossa.)

Fica claro então que a mineração de textos é mais do que uma simples busca textual. Esse conjunto de novas técnicas nos possibilita diferentes tipos de pesquisa. Podemos através da mineração de textos levantar novas questões, construir novas formas de pesquisa, elaborar novas hipóteses. Tais ferramentas permitem que diferentes metodologias coexistam na área de Humanidades, e é aí que a mineração de texto nos leva a um novo paradigma transdisciplinar²⁵.

24 Lin, “Transdisciplinarity”, 301. “Given that, text mining can be understood as an umbrella term for incorporating and implementing a wide range of tools or techniques (algorithms, methods), including data mining, machine learning, natural language processing, artificial intelligence, clustering, knowledge mining and text analysis, computational linguistics, content analysis and sentiment analysis and so forth, onto a large body of texts (usually an enormous collection of documents) to support the users’ decisions-making. Just like Lego units, there is a set of components in the field that can be assembled and reconfigured for the purposes of the tasks of the domain users.”

25 Lin, “Transdisciplinarity”, 300.

Dentre algumas ferramentas de mineração de textos podemos encontrar algumas *online* como, por exemplo, em *websites* como o Open Library²⁶, o Perseus Project²⁷, o Topos Text²⁸, o Tesserae²⁹, o Digital Dante³⁰, o eTRACES³¹, o (Dis)Similitudes³², The Valley of the Shadow³³, o Voyant Tools³⁴ e o Text-Mining the Middle Ages³⁵. Outras ferramentas de mineração de textos podem ser encontradas para *download* gratuito³⁶ ou pago na Web. Dentre elas podemos citar o Zotero³⁷, que pode contribuir de maneira muito interessante neste sentido. Para que a mineração de texto ocorra com sucesso é necessário que as ferramentas utilizadas estejam adequadas a cada projeto.

O Perseus como ferramenta inicial de mineração de textos

Discorrerei brevemente sobre a minha experiência inicial com as ferramentas de mineração de textos fornecidas pelo Perseus Project³⁸. Primeiramente, gostaria de falar um pouco sobre o que é o Perseus. Em 1985 surge a Perseus Digital Library³⁹, um projeto que foi avançando durante os anos e hoje conta com uma extensa base de dados. Segundo a própria página do Perseus, a missão maior do projeto é tornar aces-

26 Open Library, acesso em 30 nov. 2021, <https://openlibrary.org/>.

27 Perseus Project, acesso em 30 nov. 2021, <http://www.perseus.tufts.edu/hopper/>.

28 Topos Text, acesso em 30 nov. 2021, <https://topostext.org/>.

29 Tesserae, acesso em 30 nov. 2021, <http://tesserae.caset.buffalo.edu/>.

30 Digital Dante, acesso em 30 nov. 2021, <http://digitaldante.columbia.edu/>.

31 eTraces, acesso em 16 jan. 2017, <http://etraces.e-humanities.net/>.

32 Dissimilitudes, acesso em 16 jan. 2017, <http://dissimilitudes.lip6.fr:8181/#/>.

33 The Valley Project, acesso em 30 nov. 2021, <http://valley.lib.virginia.edu/>.

34 Voyant Tools, acesso em 16 jun. 2022, <https://voyant-tools.org/>.

35 Text-Mining the Middle Ages, acesso em 30 nov. 2021, <https://people.stanford.edu/widner/content/text-mining-middle-ages>.

36 “Top 26 free softwares for text analysis, text mining, text analytics”, PAT Research, acesso em 30 nov. 2021, <http://www.predictiveanalyticstoday.com/top-free-software-for-text-analysis-text-mining-text-analytics/>.

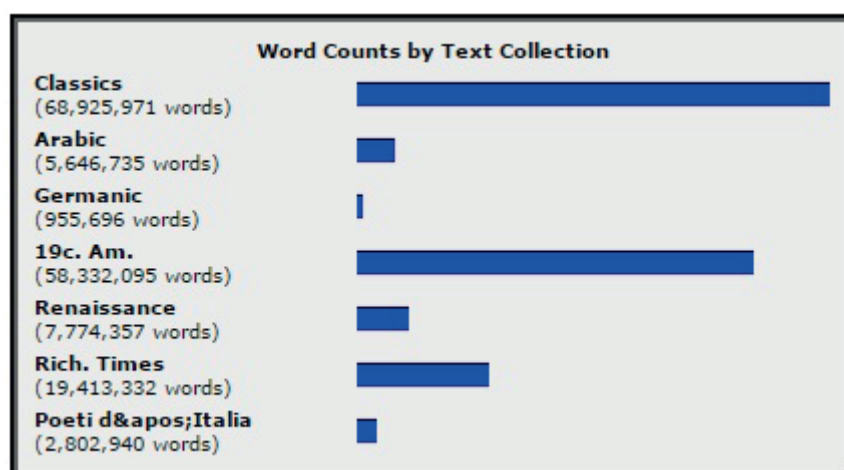
37 Zotero, acesso em 30 nov. 2021, <https://www.zotero.org/>.

38 Parte de minha experiência com as ferramentas disponibilizadas pelo Perseus Project também pode ser consultada em minha dissertação de mestrado: Ismael Wolf Ferreira, “O olhar de Tácito sobre os Brigantes: Um estudo sobre os usos do passado” (Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro – PUC-RJ, 2018).

39 Projeto Biblioteca Digital Perseus.

sível o registro “completo” da humanidade. Embora tenha um foco particular sobre o mundo greco-romano, o projeto também tem buscado contemplar outras áreas de pesquisa como a Guerra Civil Americana, por exemplo. Dentro de sua missão maior, o projeto se concentra em três categorias de acesso: *Human readable information*, *machine actionable knowledge* e *machine generated knowledge*⁴⁰⁴¹. Dentro do Perseus podemos encontrar diferentes Coleções de Textos, como podemos ver na Figura 1:

Figura 1: Contagem de palavras por Coleção de Texto⁴²



O Perseus oferece uma interessante ferramenta de busca que expõe a incidência de vezes em que os termos buscados aparecem e em quais documentos são encontrados. No caso de uma de minhas recentes pesquisas, em que trabalhei com a construção da identidade dos brigantes⁴³ através da análise dos textos de Tácito⁴⁴, procurei inicialmente

40 Informações legíveis por humanos, conhecimento acionável por máquina e conhecimento gerado pela máquina.

41 “Research”, Perseus Project, acesso em 30 nov. 2021, <http://www.perseus.tufts.edu/hopper/research>.

42 Número de palavras encontradas em cada coleção de textos presentes no Perseus.

43 Os brigantes eram um dos inúmeros povos que habitavam a Britânia, quando da chegada dos romanos a aquela ilha ainda no século I d.C.

44 Publius/Gaius Cornelius Tacitus, historiador romano que viveu entre os séculos I e II d.C.

buscar termos que já me eram familiares em inglês e sobre os quais eu já tinha consciência, através de estudos prévios, de que estes eram termos relacionados com meu tema. Realizei uma primeira busca sobre o termo *brigantes*. Neste primeiro momento tive um resultado mais vasto sobre o termo. Este foi encontrado em quatorze documentos diferentes⁴⁵, mas, apenas três destes documentos eram textos clássicos⁴⁶, todos de Cornélio Tácito. Estes eram *Historiae*⁴⁷, *De Vita Iulii Agricolae*⁴⁸ e *Annales*⁴⁹. Também no Perseus fiz a busca pelo termo *Cartismandua*⁵⁰. Este foi encontrado em seis documentos diferentes, dentre eles dois clássicos, *Annales* e *Historiae* de Tácito. Fiz ainda a busca por um terceiro termo, *Venutius*, que foi encontrado em quatro documentos diferentes, sendo dois textos clássicos, *Annales* e *Historiae* de Tácito. Busquei ainda pelo termo *Vellocatus*, tendo o encontrado em três documentos, mas apenas em um clássico, *Historiae* de Tácito. Estas buscas iniciais me permitiram uma visualização de forma ainda um tanto geral, de onde eu poderia encontrar informações iniciais sobre os brigantes, alvo primeiro de minha pesquisa. Resolvi então colocar isso em uma tabela que pudesse me auxiliar nesta visualização.

Tabela 1: A presença dos brigantes nas obras de Tácito⁵¹

	<i>Annales</i>	<i>Historiae</i>	<i>De Vita Iulii Agricolae</i>
<i>Brigantes</i>	Sim	Sim	Sim
<i>Cartismandua</i>	Sim	Sim	Não
<i>Venutius</i>	Sim	Sim	Não
<i>Vellocatus</i>	Não	Sim	Não

45 Todos em língua inglesa.

46 Ou seja, oriundos da antiguidade greco-romana.

47 *As Histórias*.

48 *Vida de Agrícola*.

49 *Anais*.

50 *Cartismandua*, *Venutius* e *Vellocatus* eram brigantes e foram citados em sua obra por Tácito.

51 Demonstra em quais obras de Tácito aparecem os termos *brigantes*, *Cartismandua*, *Venutius* e *Vellocatus*.

Após organizada esta primeira tabela, decidi que seria interessante saber quantas vezes cada um dos termos buscados aparecem nos textos de Tácito. O resultado foi este da tabela abaixo. Bastante interessante, pois permitiu que eu pudesse ter uma ideia do número de vezes em que cada termo foi citado em cada uma das obras.

Tabela 2: A presença quantitativa dos brigantes nas obras de Tácito⁵²

	<i>Annales</i>	<i>Historiae</i>	<i>De Vita Iulii Agricolae</i>
<i>Brigantes</i>	4	2	3
<i>Cartismandua</i>	3	3	0
<i>Venutius</i>	2	4	0
<i>Vellocatus</i>	0	1	0

A próxima forma de mineração de texto que procurei fazer foi através de busca direta dentro dos livros que me interessavam. É importante ressaltar que o Perseus apresenta em sua base de dados todos os textos de Tácito, tanto em latim quanto em inglês. Portanto, para um levantamento inicial eu tinha essas duas possibilidades de busca. Outros termos que me interessavam como *celts*⁵³ e *Eboracum*⁵⁴ não foram encontrados na busca dentro das obras de Tácito. Já termos em latim como *barbari*⁵⁵, por exemplo, foram encontrados de formas distintas.

⁵² Apresenta quantas vezes cada termo aparece nas obras de Tácito.

⁵³ Celtas.

⁵⁴ Atual cidade de York, no Reino Unido. Estava dentro do território que teria sido habitado inicialmente pelos brigantes.

⁵⁵ Em português: bárbaros.

Figura 2: Resultados para busca do termo barbari dentro da obra Annales de Tácito

www.perseus.tufts.edu/hopper/searchresults?target=la&inContent=true&q=barbari&doc=Perseus%3Atext%3A1999.02.0077&expand=yes

Search Results

Search ("Agamemnon", "Hom. Od. 9.1", "denarius")
All Search Options [view abbreviations]

Home Collections/Texts Perseus Catalog Research Grants Open Source About Help

Currently searching the following texts in Latin:

- Cornelius Tacitus, *Annales* (ed. Charles Dennis Fisher)

Showing 1 - 1 of 1 document results in Latin.

Cornelius Tacitus, *Annales* [Less](#)
(Latin) (English, ed. Alfred John Church, William Jackson Brodribb)

book 1, chapter 57: ... quis circumsedebatur, validiore apud eos Arminio quoniam bellum suadebat: nam **barbaris**, quanto quis audacia promptus, tanto magis fidus rebusque motis

book 1, chapter 61: ... telorum equorumque artus, simul truncis arborum antefixa ora. lucis propinquis **barbarae** arae, apud quas tribunos ac primorum ordinum centuriones mactaverant.

book 1, chapter 64: **Barbari** perfringere stationes seque inferre munitoribus nisi lacesunt, circumgrediuntur, occurrant:

book 1, chapter 65: Nox per diversa iniquis, cum **barbari** festis epulis, laeto cantu aut truci sonore subiecta vallium

book 1, chapter 68: ... per umida et impedita circumvenirent suadente, atrociora Inguimero et laeta **barbaris**, ut vallum armis ambient: promptam expugnationem, plures captivos, incorruptam

book 2, chapter 2: ... accirent. magnificum id sibi creditit Caesar auxitque opibus. et accipere **barbari** laetantes, ut ferme ad nova imperia. mox subit pudor

book 2, chapter 14: ... si ratio adsit, silvas et saltus; nec enim immensa **barbarorum** scuta, enormis hastas inter truncos arborum et enata humo

book 2, chapter 16: ... et pura humo inter arborum truncos. campum et prima silvarum **barbara** acies tenuit: soli Cherusci iuga insedere ut proeliantibus Romanis

book 2, chapter 21: ... miles, cui scutum pectori adpressum et insidens capulo manus, latos **barbarorum** artus, nuda ora foderet viamque strage hostium aperiret, inprompto

book 2, chapter 56: ... instituta et cultum Armeniorum aemulatus, venatu epulis et quae alia **barbari** celebrant, proceres plebemque iuxta devinerat. igitur Germanicus in urbe

book 2, chapter 63: ... et Vibilio duce receptusque, Forum Iulium, Narbonensis Galliae coloniam, mittitur. **barbari** utrumque comitati, ne quietas provincias immixti turbarent, Danuvium ultra

book 2, chapter 88: ... septem et triginta annos vitae, duodecim potentiae explevit, caniturque adhuc **barbaras** apud gentis, Graecorum annalibus ignotus, qui sua tantum mirantur,

book 2, chapter 91: ... sedesque in tres partes abduxit. **barbari** filii, inchois foronibus Democritus invenit

Refine This Search [hide](#)

Language: Latin

Required words: barbari Expand

Required phrase:

Allowed words: Expand

Excluded words: Expand

Refine search
(This searches within the currently selected documents. To search within all documents, use the form below.)

All Matching Documents (1) [show](#)

Matching Lemmas (2) [hide](#)

- barbarum: "a plaster applied to raw wounds" (entry in Lewis & Short)
- barbarus: "of strange speech, speaking jargon, unintelligible;" (entry in Lewis & Short Elem. Lewis)

O termo em latim *barbari* foi encontrado cinquenta e uma vezes dentro da obra *Annales*. Como observamos na Figura 2, o Perseus oferece uma visualização em que são expostas as referências dos locais onde foram encontrados os termos. Isso pode contribuir para um mapeamento inicial dos termos que buscamos. Pensar isso como uma ferramenta que ajuda a mapear certos termos que procuramos é algo extremamente interessante e que pode contribuir significativamente para o início de uma pesquisa. É importante que se perceba que o Perseus apresentou resultados não apenas para o termo conforme foi escrito na

busca, mas também algumas derivações com terminações diferentes como *barbaris*, *barbarorum*, *barbara* e *barbaras*. Uma demonstração de que o mecanismo de pesquisa realizou a tarefa prevendo estas pequenas variações linguísticas.

Embora minha utilização do Perseus, enquanto ferramenta para mineração de texto, tenha tido um caráter mais quantitativo, é importante dizer que existem outras possibilidades oferecidas pelo mesmo e que podem levar a uma análise qualitativa. Além desses mecanismos de mineração de textos, o Perseus também oferece dicionários que podem auxiliar o pesquisador. Durante a leitura de um determinado texto basta apenas um clique em cima da palavra desejada para que você possa ler as definições do dicionário.

Mesmo tendo tido uma experiência inicial interessante com essa ferramenta, é preciso expor o fato de que existem algumas dificuldades e pontos a melhorar. Uma das dificuldades que encontrei na utilização do Perseus foi o fato de ele não permitir *download* de seus textos. Todas as consultas só podem ser realizadas *online*, o que impede que você exporte o texto para algum *software* que realize outra forma de mineração de textos.

Considerações finais

A mineração de textos tem se mostrado útil para diversos pesquisadores de diferentes áreas. Basta uma simples busca em bases de dados acadêmicas para chegarmos a esta fácil constatação. Nas Humanidades Digitais isto não tem sido diferente e os trabalhos de pesquisadores como, por exemplo, Robert K. Nelson⁵⁶ e Ruby Mendenhall⁵⁷ comprovam isso. Além do mais, propostas como a do Perseus Project reforçam a ideia de que um movimento para a criação e divulgação de bases de dados abertos para a comunidade está cada vez mais forte. A palavra de ordem é abrir os dados e colaborar. Podemos perceber que a História Digital compartilha da mesma perspectiva da chamada História

⁵⁶ “Mining the Dispatch”, acesso em 22 jun. 2022, <https://dsl.richmond.edu/dispatch/>.

⁵⁷ “Rescued History”, acesso em 22 jun. 2022, <https://beta.nsf.gov/news/rescued-history>.

Pública. Neste sentido busca-se um maior diálogo com/entre as comunidades envolvidas.

Assim sendo, faz-se necessária uma colaboração cada vez maior entre as humanidades e as novas tecnologias. Diferentes tecnologias e profissionais de diferentes áreas devem atuar de forma conjunta. Possibilidades de apresentar e analisar de maneira rápida e eficiente uma enorme quantidade de dados são, atualmente, uma realidade para as Humanidades Digitais. As novas ferramentas nos permitem que analisemos antigas fontes com novas abordagens, permitindo novos questionamentos e novas hipóteses. Portanto, os historiadores precisam de estar a par das possibilidades e das ferramentas disponíveis na contemporaneidade, sabendo escolher o que for melhor de acordo com as condições e as exigências de suas pesquisas. Neste sentido, o Perseus Project apresenta ferramentas que, se usadas de forma consciente, podem ser bastante úteis para os historiadores.

REFERÊNCIAS BIBLIOGRÁFICAS

Ambrósio, Ana Paula L., e Edilson Andrade Martins Morais. “Mineração de textos”. Instituto de Informática da UFG, acesso em 30 nov. 2021, http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf.

Arnold, Bettina, Kevin Garstki, e Matthew L. Murray. “Reconstituting Community: 3D Visualization and Early Iron Age Social Organization in the Heuneburg Mortuary Landscape”. *Journal of Archaeological Science* 54 (2015): 23-30.

Berry, David M. “Introduction”. In *Understanding Digital Humanities*, organizado por David M. Berry, 1-20. Basingstoke: Palgrave Macmillan, 2012.

Brier, Steve. “Historians and Hypertext: Is It More Than Hype?”. In *Clio Wired: The Future of The Past in The Digital Age*, organizado por Roy Rosenzweig, 85-91. Nova Iorque: Columbia University Press, 2011.

Evans, Leighton, e Sian Rees. “An Interpretation of Digital Humanities”. In *Understanding Digital Humanities*, organizado por David Berry, 21-41. Basingstoke: Palgrave Macmillan, 2012.

Ferreira, Ismael Wolf. “O olhar de Tácito sobre os Brigantes: um estudo sobre os usos do passado”. Dissertação de mestrado, Pontifícia Universidade Católica do Rio de Janeiro – PUC-RJ, 2018.

Hearst, Marti. “What Is Text Mining?”, SIMS, UC Berkeley, acesso em 30 nov. 2021, <http://people.ischool.berkeley.edu/~hearst/text-mining.html>.

Lin, Yu-wei. “Transdisciplinarity and Digital Humanities: Lesson Learned from Developing Text-Mining Tools for Textual Analysis”. In *Understanding Digital Humanities*, organizado por David M. Berry, 295-314. Basingstoke: Palgrave Macmillan, 2012.

Middleton, Roger, e Peter Wardley. “Information Technology in Economic and Social History: The Computer as Philosopher’s Stone or Pandora’s Box?”. *The Economic History Review* 43, n.º 4 (1990): 667-96.

Priego, Laia Pujol, Jonathan Wareham, e Angelo Kenneth S. Romasanta. “The Puzzle of Sharing Scientific Data”. *Industry and Innovation* 29, n.º 2 (2022): 219-50.

Terrel, John Edward. “Social Network Analysis and the Practice of History”. In *Network Analysis in Archaeology: New Approaches to Regional Interaction*, organizado por Carl Knappett, 17-42. Oxford: Oxford University Press, 2013.

Thwaites, Harold. “Digital Heritage: What Happens When We Digitize Everything?”. In *Visual Heritage in the Digital Age*, organizado por Henry Chapman, Eugene Ch’ng, Vincent Gaffney, 327-348. Londres: Springer-Verlag, 2013.

UK JISC-funded National Centre for Text Mining (NaCTeM). “Text Mining”. Acesso em 16 jan. 2017. <http://www.jisc.ac.uk/publications/briefingpapers/2008/bptextminingv2.aspx>.

Referência para citação:

Wolf, Ismael. “História digital, mineração de texto e análise textual: algumas possibilidades através da utilização de ferramentas oferecidas pelo Perseus Project”. *Práticas da História, Journal on Theory, Historiography and Uses of the Past*, n.º 14 (2022): 151-168.