

**HUMANIDADES DIGITAIS E A SIMBIOSE ENTRE HUMANO E MÁQUINA:
ALGUMAS REFLEXÕES COMPARATIVAS ENTRE A INTERPRETAÇÃO E A
MINERAÇÃO DE TEXTOS**

José Claudio Matos

Universidade Federal de Santa Catarina
doutortodd@gmail.com

Eliana Maria dos Santos Bahia Jacinto

Universidade Federal de Santa Catarina
elianambahia@gmail.com

Edgar Bisset Alvarez

Universidade Federal de Santa Catarina
ebicet@gmail.com

Resumo

Examina a relação entre a leitura e os programas computacionais de mineração de textos, no contexto das Humanidades Digitais (HD). Através da pesquisa bibliográfica e da reunião de fontes textuais relevantes para o argumento, esta relação é descrita sob a forma de uma simbiose entre o ser humano e o computador, nos termos propostos inicialmente por Asimov (2005). Uma concepção da leitura como operação da inteligência é posta em diálogo com o conceito e as aplicações da mineração de textos. Da reflexão sobre a simbiose entre leitura e mineração de textos é obtido, como resultado, um ponto de vista em relação às HD, entendida como um conjunto de estratégias para o crescimento e a adaptação da cultura humana no ambiente digital.

Palavras-chave: Humanidades Digitais. Leitura. Mineração de Textos. Inteligência artificial.

**DIGITAL HUMANITIES AND THE HUMAN-MACHINE SYMBIOSIS:
SOME COMPARATIVE REFLECTIONS BETWEEN INTERPRETATION AND
TEXT MINING**

Abstract

It examines the relationship between reading and computational programs of text mining in the context of Digital Humanities (HD). Through the bibliographic research and the reunion of textual sources relevant to the argument, this relationship is described in the form of a symbiosis between the human and the computer, in terms initially proposed by Asimov (2005). A conception of reading as an operation of intelligence is put into dialogue with the concept and applications of text mining. From the reflection on the symbiosis between reading and text mining is obtained, as a result, a point of view in respect to the HD, understood as a set of strategies for the growth and adaptation of human culture in the digital environment.

Keywords: Digital Humanities. Reading. Text Mining. Artificial Intelligence.



1 PREÂMBULO: LITERATURA, HUMANIDADES E ANTECIPAÇÃO TECNOLÓGICA

O conhecido ensaísta e escritor Isaac Asimov é autor do prefácio de uma antologia de contos de ficção científica, publicada no Brasil com o título de *Histórias de Robôs* (WARRICK; GREENBERG, 2005). Na edição original, de 1983, o título era mais sugestivo: *Machines that Think – Máquinas que Pensam*. O texto de Asimov, ao invés de simplesmente comentar algo das histórias ali reunidas, discute em forma de ensaio o interessante problema que o autor define como “medo irracional que o homem tem dos autômatos” (ASIMOV, 2005, p. 7). Trata, em especial, do medo da crescente capacidade dos computadores para realizar funções complexas, de forma extremamente rápida e inteligente.

Segundo Asimov: “Temos que admitir que, pelo menos como concepção, o medo não deixa de ser justificado. Não há nenhum limite teórico visível para a complexidade e ‘inteligência’ do computador” (ASIMOV, 2005, p. 13). Ele dirige seu argumento a uma conclusão conciliadora, em termos de que a inteligência humana e a inteligência artificial evoluíram por processos diferentes, apresentam melhor desempenho em funções bastante diferentes e podem coevoluir de forma interativa, conforme os mais elevados objetivos e interesses humanos.

O que Asimov supõe, considerando o estado de desenvolvimento tecnológico do final do século XX, é que

[...] duas inteligências diferentes, especializando-se em objetivos diversos, cada qual com sua utilidade, podem, num relacionamento simbiótico, aprender a colaborar com a lei natural do Universo de forma mais eficiente do que separadamente (ASIMOV, 2005, p. 15).

Esta passagem incorpora, em grande medida, o sentido do problema que o presente artigo espera discutir. Trata-se da indagação sobre quais fundamentos permitem propor a ideia de uma relação simbiótica entre a mente humana e a inteligência artificial, cujo resultado seja benéfico para o desenvolvimento da inteligência e da cultura em geral – admitindo a hipótese de que é isso que Asimov quis dizer com “lei geral do Universo”.

Com sutileza e habilidade, Asimov evita caracterizar qualquer conflito de interesses entre seres humanos e computadores, assim como qualquer sobreposição de funções. Seu discurso é veemente ao defender que estes estão a serviço da humanidade, como instrumentos ou, no máximo, como ajudantes. Mesmo assim é preciso reconhecer os computadores como artefatos complexos, capazes de operações inteligentes e - especialmente - como os

constituintes do novo e penetrante ambiente informacional que hoje em dia chamamos de ‘ciberespaço’ ou ‘infosfera’.

O escritor menciona a agenda de “questões que implicam em conceitos sobre o que está certo ou errado” (ASIMOV, 2005, p. 17), e relaciona o futuro da tecnologia digital com o conjunto de princípios, valores, conhecimentos e formas estéticas que caracterizam a cultura humana de forma geral. Essa tonalidade humanística de seu ensaio, a profundidade das questões suscitadas por sua leitura, se incorporada criticamente pelo leitor, enriquecem a experiência provocada pelos contos da antologia da qual seu ensaio é o prefácio.

Para os fins da presente discussão, a ideia de *simbiose* entre a mente humana e o computador será discutida, no contexto de uma atividade extremamente familiar e de alta importância cognitiva e social, que é a leitura ou interpretação de textos escritos. Além disso, o cultivo de valores humanísticos como a liberdade de expressão e o acesso ao conhecimento constitui o cenário deste movimento investigativo, na forma da tendência contemporânea de pensamento e trabalho intelectual conhecida como Humanidades Digitais (HD) - *Digital Humanities*.

Em um artigo intitulado “Humanidades digitais, dejalas ser”, Juan Manuel Cuartas-Restrepo aponta para uma mudança operada na própria atividade de pesquisa e investigação pelo advento da cultura digital. Segundo ele, a própria escrita no campo das HD, difere das formas tradicionais de comunicação do conhecimento. Afirma ele que

[...] em essência, o que mudou nas ‘humanidades digitais’ é a escritura, tão sujeita à razão e à lógica das linguagens verbais, mas tão diametralmente diferente do duplo movimento de abstração e descrição de outras formas de representação como as gráficas e as imagens¹ (CUARTAS, 2017, p. 73, tradução nossa).

Com esta passagem, Cuartas indica a necessidade e possibilidade da multiplicação de formas de comunicação, a fim de que se promova o ideal humanístico do acesso à cultura e ao conhecimento, no movimento evolutivo em que se inserem as HD. A referência ao pensamento de Isaac Asimov, e especialmente ao seu conceito de ‘relacionamento simbiótico’ entre o humano e o computador é tomada, no presente trabalho, como parte desta diversificação da escrita, e desse “deslocamento do centro de operações da vida acadêmica²” (CUARTAS, 2017, p. 73, tradução nossa), que segundo este estudioso é um traço marcante das HD.

Estas considerações iniciais tiveram o objetivo de situar o leitor no campo de discussão em que este trabalho pretende se movimentar. Fizeram isso indicando que a inspiração ou a

¹ “[...] en esencia lo que ha cambiado en las “humanidades digitales” es la escritura, tan sujeta a la razón y a la lógica de los lenguajes verbales, pero tan diametralmente diferente al doble movimiento de abstracción y descripción de otras formas de representación como las gráficas y las imágenes” (CUARTAS, 2017, p. 73).

² “[...] desplazamiento del centro de operaciones de la vida académica” (CUARTAS, 2017, p. 73).

origem da questão a ser tratada se encontra num texto de caráter literário, embora motivado por temas científicos. Esta modalidade da experiência, que combina uma cultura artística e literária com uma cultura científica e tecnológica, parece ser um exemplo interessante da espécie de mentalidade acolhedora, ampla e inclusiva encontrada nos projetos e estudos das HD. Isaac Asimov, mesmo sendo escritor de ficção científica, antecipou no mencionado ensaio - como também em suas histórias - o importante tema da simbiose entre a mente e o computador, que aqui se espera explorar, particularmente para o caso da interpretação e leitura de textos escritos.

Feitas estas considerações, segue-se a contextualização e caracterização do estudo, o exame das informações e fontes bibliográficas em vista dos objetivos, e a exposição e análise dos resultados obtidos.

2 INTRODUÇÃO

As HD vêm se constituindo como um campo de iniciativas e de discussões em torno da relação entre a cultura humanística tradicional, de um lado, e a cultura marcada pelas tecnologias digitais, de outro. Priani formula de maneira bastante adequada sua descrição das HD como “espaço de aplicação e reflexão acadêmica sobre o uso de métodos e sistemas computacionais nas humanidades³” (PRIANI, 2015, p. 1218, tradução nossa). Mais do que um campo disciplinar específico, é proveitoso procurar compreender as HD como um conjunto de valores e de atitudes promovidos a partir das investigações e iniciativas práticas que se baseiam nesta relação entre as humanidades e a tecnologia digital.

Em um artigo intitulado “Beyond the tale of the ‘two cultures’: filling the gap between algorithms and interpretation” (TĂUT, 2017), o estudioso Val Codrin Tăut argumenta em favor de um significado compreensivo e incluyente das HD. Sua proposta é de um “entendimento não-dogmático das Humanidades Digitais⁴” (TĂUT, 2017, p. 2, tradução nossa), a partir do qual se possa integrar a atividade de interpretação e apreciação do legado cultural, com a computação e os processos algorítmicos da tecnologia digital. Conforme será apresentado adiante, esta integração entre a interpretação e a computação é um processo que não pode se desenvolver sem a formulação de um aparato conceitual, e sem a proposição experimental de variadas metodologias e formas de aprendizagem. Neste horizonte conceitual, a relação entre a cultura humanística e a cultura digital se revela bastante complexa e multifacetada. Um dos

³ “[...] *espacio de aplicación y reflexión académica sobre el uso de métodos y sistemas computacionales en las humanidades*” (PRIANI, 2015, p. 1218).

⁴ “[...] *non-dogmatic understanding of Digital Humanities*” (TĂUT, 2017, p. 2).

casos desta diversidade de aspectos é o da relação sempre controversa entre o desempenho da inteligência humana e o desempenho dos computadores digitais.

Desde sua invenção, as máquinas que processam informação suscitaram questionamentos acerca da sobreposição ou da concorrência com aquilo que a mente humana faz. Parte desses questionamentos manifesta a ideia de que a chamada inteligência artificial possa ser entendida como uma equivalente da inteligência humana que a projetou. Outra parte chega a expressar certo temor de que a inteligência artificial possa substituir de forma indesejável o desempenho da mente em tarefas que exigem raciocínio e capacidade de processamento de informações. Ora, as HD, pelo que se observa pelos inúmeros projetos desenvolvidos neste campo, vêm concentrando sua atenção na possibilidade de aperfeiçoar o emprego da inteligência artificial a fim de promover valores práticos e teóricos como a diversidade da cultura humana e a importância da educação e do esclarecimento, pela via do acesso autônomo e crítico às realizações desta cultura.

No trabalho intitulado “La diversidad cultural em las HD” (SÁNCHEZ, 2018), a autora discute a integração indissociável entre o emprego de procedimentos de pesquisa e investigação com os princípios humanísticos. Exemplos desta integração envolvem a valorização das diversas culturas e a garantia de amplo acesso e inclusão digital das pessoas de todas as sociedades. Interessante em seu argumento é o destaque para “a línguas e seus usos, o resgate do patrimônio tangível e intangível e os sistemas educativos⁵” (SÁNCHEZ, 2018, p. 183, tradução nossa). Sánchez reforça a noção de que a linguagem e a comunicação são os meios de expressão da cultura, e que qualquer iniciativa para desenvolver uma cultura digital da diversidade precisa dar conta da diversidade de línguas, do acesso ao legado da cultura escrita, e ao aprendizado da escrita e da leitura. Tăut, acima citado, reforça este ponto de vista em seu estudo, ao afirmar que “o primeiro passo no desenvolvimento das HD foi largamente dominado pela análise linguística e literária⁶” (TĂUT, 2017, p. 2, tradução nossa).

O que o presente trabalho pretende acrescentar a estas reflexões é a percepção de que, na atualidade, um dos instrumentos de investigação e promoção das HD são as ferramentas digitais conhecidas como “descoberta de conhecimento em textos”, ou “mineração de textos”. O que se espera discutir são as questões que nascem da correlação entre a leitura, como um processo humano de interpretação, e a mineração de textos, como um processo algorítmico de

⁵ “[...] *la lengua y sus usos, el rescate Del patrimonio tangible e intangible y los sistemas educativos*” (SÁNCHEZ, 2018, p. 183).

⁶ “[...] *this first step in the development of Digital Humanities was largely dominated by linguistic and literary analysis*” (TĂUT, 2017, p. 2).

computação. A hipótese que se espera submeter ao exame é a de que a *simbiose* entre a mente humana e o computador - ou especificamente entre a interpretação e a mineração de textos - expressa a forma mais bem fundamentada e mais promissora desta correlação, no horizonte das HD.

A seguir, o estudo apresenta seus aspectos metodológicos, para em seguida passar a compor uma concepção teórica da leitura ou interpretação, assim como da mineração de textos, com base na discussão especializada, revelada pela pesquisa de fontes bibliográficas.

3 PROCEDIMENTOS METODOLÓGICOS

Este artigo resultou de um processo de reflexão que buscou formular sua questão principal a partir de um histórico de leituras dispersas ao longo de anos de atividades de pesquisa e ensino, nos campos da filosofia, educação, ciência da informação, ciências cognitivas e literatura. Este percurso mais individual de estudos, nem por isso se ressentiu diante da exigência de evidências e justificativas que possam ser alvo de uma abordagem crítica e objetiva por parte do leitor.

Dando atenção a certa diversificação metodológica, na esteira do que acima foi comentado, quando feita menção ao trabalho de Cuartas (2017), o formato deste documento expressa um percurso intelectual que incorpora elementos culturais e teóricos advindos de uma vivência de profunda reflexão sobre os assuntos aqui discutidos. Em geral tomados em separado, pela primeira vez se empreende neste trabalho a tentativa de construir a ponte entre temas da cultura literária e humanística, com temas científicos e tecnológicos inerentes às HD. O tema da leitura foi o agente catalisador dos conhecimentos derivados dessas fontes, e seu tratamento dentro do cenário inovador e flexível das HD exigiu, naturalmente, o reforço de uma pesquisa bibliográfica mais rigorosa e objetiva.

Embora não se trate exatamente do relatório de um projeto de pesquisa científica em sentido estrito, este estudo segue o procedimento da pesquisa bibliográfica, concentrada em artigos de periódicos científicos qualificados. As bases consultadas foram o Portal de Periódicos CAPES e a *Journal Storage* (JSTOR). Os termos da busca foram “leitura”, “humanidades digitais”, “mineração de textos” e algumas combinações destes termos segundo os operadores booleanos para recuperação da informação. A seleção das fontes foi operada pela leitura dos resumos dos artigos recuperados, com a finalidade de identificar aqueles que poderiam contribuir para a discussão do tema central do artigo.

Uma das vantagens de se praticar a reflexão em HD, é a de absorver a acessibilidade informacional do ciberespaço como um constante recurso do debate e da reflexão. Assim considerada, a base bibliográfica do presente artigo é sempre uma sugestão à participação do leitor como um interagente crítico e um interlocutor ativo na discussão que se propõe. Isso não é uma justificativa para o desleixo, a informalidade ou a ausência de critérios de qualidade da pesquisa, mas sim um chamado a uma forma de ler e consumir informação científica, em que outros hábitos da inteligência e outras formas de atestar a legitimidade dos conhecimentos possam ser considerados.

Segue-se aqui uma caracterização dos conceitos de leitura e de mineração de textos, fundamentada nestas leituras dispersas, em combinação com as fontes resultantes da pesquisa bibliográfica por artigos especializados, com o objetivo de explorar as relações entre ambos os processos e defender a ideia de uma relação simbiótica entre o humano e a máquina, no âmbito das HD.

4 SOBRE LEITURA E INTERPRETAÇÃO

A leitura é uma atividade tão familiar e habitual, que muitas vezes é necessário certo esforço de abstração para que se possam considerar teoricamente as questões que a envolvem. Até aqui se optou por fazer referência a ela combinando os termos “leitura” e “interpretação”. Quando se diz “leitura”, o termo costuma significar o processo de reconhecimento ou construção do significado de textos escritos em linguagem natural – como o português ou o inglês. Mas, para acentuar o aspecto ativo e a profundidade cognitiva desta operação, se acrescenta à “leitura” o complemento “interpretação”. A partir daqui será empregada somente a palavra “leitura”, para referir à atividade da mente humana quando decifra, codifica ou estabelece o significado de um segmento de texto escrito.

Além disso, como observa Elsa Ramírez Leyva, em seu artigo “¿Qué es leer? ¿Qué es la lectura?” (LEVYA, 2009), desde a explosão informacional a leitura se tornou uma questão de interesse nas ciências sociais. Conforme a autora, os estudiosos da leitura “começaram a questionar seus paradigmas, assim como um campo de interesse para o âmbito cultural, em um contexto donde se vislumbravam mudanças do modelo cultural, do projeto social e da tecnologia⁷” (LEVYA, 2009, p. 163, tradução nossa). Todos os aspectos do fenômeno da

⁷ “[...] empezaron a cuestionar sus paradigmas, sino también un campo de interés para el ámbito cultural, en un contexto donde se vislumbraban cambios del modelo cultural, el proyecto social y la tecnología” (LEVYA, 2009, p. 163).

linguagem e da comunicação tiveram um aumento de relevância nos estudos teóricos, a partir da constituição cada vez mais evidente de uma cultura virtual, e de uma forma de vida social marcada pelo fluxo intenso de mensagens. Assim, “a leitura e a formação de leitores⁸” (LEVYA, 2009, p. 163, tradução nossa), especialmente a chamada leitura crítica ou reflexiva, se evidencia como um aspecto fundamental da vida social na cultura contemporânea.

Para fins da presente discussão será dada ênfase ao aspecto cognitivo da leitura de textos escritos: leitura como uma operação mental de dar significado a um texto escrito. Acerca disso, segue-se a recomendação de José Morais, em seu livro *A arte de ler* (2002). Em suas palavras:

Para compreender o que é a leitura temos que evitar estender o campo de aplicação do nosso objeto de estudo. Aumentando a extensão do conceito, alguns pensam certamente em aumentar sua importância. Dessa maneira, na verdade, não se saberia mais o que exatamente se estuda (MORAIS, 2002, p. 15).

Reconhecendo a relevância social e cultural desta atividade se espera, ainda assim, destacar o entendimento da leitura como uma operação ativa de construção de significado, realizada por uma pessoa, no contato com uma estrutura simbólica que é o texto escrito. Em outra ocasião, pesquisas relacionadas com este problema resultaram na seguinte formulação:

O termo “leitura” está sendo usado no sentido de um tipo bem específico de “interpretação”, ou seja, esforço intelectual de atribuição de significado a objetos que atuam como signos, ou símbolos, com o objetivo de compreender e, portanto, ser capaz de fazer uso destes objetos. Quando se diz “leitura”, não se está utilizando o termo em sentido amplo e genérico, mas no sentido da interpretação dos símbolos linguísticos componentes de um texto, estruturados de forma a veicular um significado compreensível pelo leitor (MATOS, 2013, p. 582).

133

Esta demarcação liberta o argumento da exigência de dar conta de todas as muitas questões periféricas em torno da leitura, a fim de enfatizar a operação pessoal da interpretação de textos.

Finalmente, a leitura é considerada uma operação da mente, uma das marcas características da inteligência e da comunicação humana. Reconstruir um pensamento através da decifração e da manipulação de símbolos escritos é uma das operações mais fundamentais para o tipo de mente que a espécie humana desenvolveu. A leitura de textos escritos equipa a mente individual com os conhecimentos e hábitos de inteligência necessários a uma vida plena e significativa, nos ambientes sociais complexos. Nos primórdios da modernidade, o velho Descartes dizia que

[...] a leitura de todos os bons livros é qual uma *conversação* com as pessoas mais qualificadas dos séculos passados, que foram seus autores, e até uma conversação premeditada, na qual eles nos revelam tão-somente os melhores de seus pensamentos (DESCARTES, 1984, p. 17, grifo nosso).

⁸ “[...] *la lectura y la formación de lectores*” (LEVYA, 2009, p. 163).

Esta aproximação da leitura com a conversa é bastante difundida. A leitura está tão profundamente atrelada ao comportamento tipicamente humano, que autoriza que seja identificada - metaforicamente - com a conversação.

Grande parte do ideal humanístico de esclarecimento e emancipação, que se observa nas bases do pensamento moderno, está apoiada na difusão da leitura e da escrita, especialmente a partir da revolução causada pela tecnologia de impressão. Ler indica um traço humano característico, na medida em que envolve a capacidade de reconhecer o significado do texto, a ponto de tornar possível acreditar ou duvidar, questionar e refletir e, em geral, gerar respostas comportamentais e emocionais a partir do texto.

Recentemente observam-se programas computacionais realizando a decodificação de textos, geralmente operando com um volume de dados muito acima da capacidade de um leitor humano. O desempenho de tais programas inevitavelmente desperta interesse, em vista das inúmeras aplicações possíveis e da similaridade, ou distinção, em relação à leitura realizada pelas pessoas.

5 SOBRE MINERAÇÃO DE TEXTOS

134

A evolução da computação digital permitiu o desenvolvimento de programas capazes de obter informações relevantes, analisando volumes cada vez maiores de dados. Na primeira década deste século, os processos de mineração de textos começaram a se tornar conhecidos da comunidade científica. Em um relatório técnico de 2007, dedicado a caracterizar os procedimentos de mineração de textos, Morais e Ambrósio fornecem a seguinte definição:

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, que normalmente não poderiam ser recuperadas utilizando métodos tradicionais de consulta, pois a informação contida nestes textos não pode ser obtida de forma direta (MORAIS & AMBRÓSIO, 2007, p.1).

Esta definição evidencia que o processo de mineração de textos é realizado de forma automática por programas computacionais, que detectam padrões e singularidades na estrutura de textos escritos por pessoas. Aranha e Passos, em seu artigo, reforçam esta forma de conceber, ao afirmarem que “Mineração de textos consiste em extrair regularidades, padrões ou tendências de grandes volumes de textos em linguagem natural” (ARANHA; PASSOS, 2006, p. 1). Faltaria talvez acrescentar que alguns desses programas são capazes de estabelecer

ordenações e identificar relações entre conceitos nos textos e classificando-os conforme qualquer critério desejado, segundo sua estrutura.

Perceba-se que a mineração de textos é um ramo específico do campo mais amplo conhecido como “mineração de dados”, que em geral significa “processo de extração ou mineração de conhecimento a partir de grandes volumes de dados” (MORAIS; AMBRÓSIO, 2007, p. 4). Estes dados geralmente se encontram organizados e preparados para receberem tratamento algorítmico, o que não acontece no caso da mineração de textos, cuja estrutura segue a forma da linguagem natural, visando a comunicação humana. Textos são em geral escritos por pessoas, a fim de serem lidos por outras pessoas. Os algoritmos de mineração de textos, portanto, são especiais porque computam informações em estruturas de dados que não foram construídas para se oferecer a este propósito.

Além disso, os autores do referido relatório relacionam a mineração de textos com a noção de “processo de descoberta de conhecimento” (MORAIS; AMBRÓSIO, 2007). Seria preciso tomar cuidado com a forma dessa expressão, na medida em que faz sentido questionar: *quem* faz a descoberta, e este conhecimento é conhecimento para *quem*? O texto é algo feito para ser ‘descoberto’ pela interpretação de um humano falante da língua. Quando se fala em “descoberta de conhecimento”, o sentido literal da expressão seria algo como: estabelecimento de padrões não triviais de ocorrência de termos e formas sintáticas estruturadas logicamente, com propósitos de análise estatística de grande volume de textos, e por técnicas sintáticas baseadas em procedimentos automáticos. Claro está que quem faz a descoberta de conhecimento é o usuário, tendo o programa computacional como seu instrumento.

135

As aplicações da mineração de textos envolvem a reunião de informações de textos escritos em redes sociais ou postagens na internet, para uso em decisões organizacionais. No ambiente acadêmico, ela se aproxima da chamada analítica cultural - *cultural analytics*, que é definida por Manovich como “a análise de conjuntos massivos de dados e fluxos culturais usando técnicas de computação e visualização⁹” (MANOVICHI, 2016, p. 1). Já foi mencionado que o texto em linguagem natural constitui um tipo especial de estrutura de dados, pois sua estrutura tem como propósito a comunicação de significados entre pessoas.

Aplicações especialmente interessantes da mineração de textos são aquelas que se desenvolvem em apoio ao processo tipicamente humano de leitura. Diversas produções apresentam resultados do emprego da mineração de textos na aprendizagem e na sofisticação

⁹ “[...] *the analysis of massive cultural data sets and flows using computational and visualization techniques*” (MANOVICHI, 2016, p. 1).

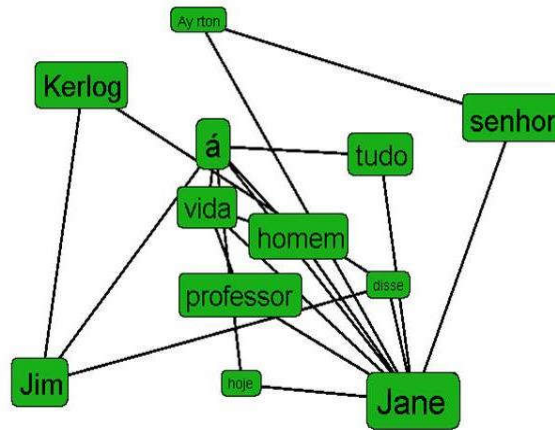
de processos de leitura e de escrita de textos. Macedo *et al.* (2011), descrevem estratégias para a aplicação da mineração de textos a fim de “levar o estudante a refletir tanto sobre os textos individualmente, quanto sobre seus relacionamentos com outros” (p. 1019). Muciolo (2017) é ainda mais ousado ao se referir à relação entre leitura e tecnologia digital. Cita o “Gutenberg Project” como um antecedente das recentes aplicações desta tecnologia que, segundo o autor, permite “proporcionar conhecimento do material textual que ultrapassa a barreira do conhecimento adquirido pela leitura” (p. 6). Parece haver relativo consenso acerca de que a mineração colabora com a leitura, permitindo um acréscimo de significado aos textos.

Tanto no sentido da aprendizagem de formas mais reflexivas e aprofundadas de leitura, quanto do ponto de vista da facilidade em manejar montantes muito grandes de textos, com vistas à organização e sistematização, a mineração de textos tem apresentado resultados promissores, a julgar pelo que os estudos vêm apresentando. O comportamento de computadores digitais que analisam textos e favorecem a “descoberta de conhecimento” parece muito mais interessante quando é descrito em termos mentais. A capacidade de um programa, de fornecer um gráfico com os termos de maior ocorrência em um texto, e a estrutura das relações entre os termos, por exemplo, tem toda a aparência de um comportamento organizado e inteligente.

Não é o caso aqui de fazer a análise de cada um dos programas de mineração de texto disponíveis, embora alguns dos mais conhecidos possam ser mencionados como exemplo. É o caso do Sobek, um algoritmo de mineração de textos criado em 2007, que produz gráficos com os termos mais relevantes e as relações entre os principais termos de um texto em linguagem natural. Segundo o tutorial desenvolvido pelo Grupo de Pesquisa Gtech.Edu, da UFRGS, o Sobek, por sua interface de fácil utilização, geralmente é adotado como ferramenta em projetos de apoio à leitura crítica e ao letramento, mesmo no caso de crianças em idade escolar (UFRGS, 2017).

A figura 1 apresenta o grafo produzido pelo programa Sobek, a partir do texto do romance *O Presidente Negro*, de Monteiro Lobato (2009).

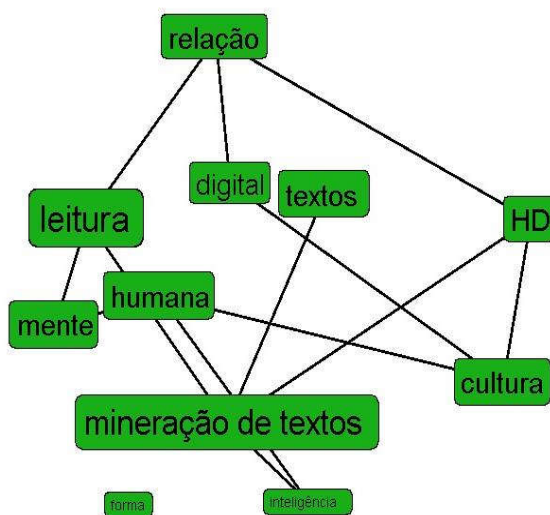
Figura 1 – Grafo de mineração do texto de *O Presidente Negro*



Fonte: Elaborado pelos autores utilizando o programa Sobek (2019)

Pode-se perceber que o grafo apresenta termos recorrentes no texto escrito por Monteiro Lobato, dando destaque diferente conforme a frequência com que o termo ocorre. As linhas representam relações textuais entre os termos. Tanto os balões diferenciados que representam a frequência dos termos, como as linhas que representam relações, são indicativos do reconhecimento de padrões no texto por parte do programa. Por se tratar de uma narrativa, é de esperar que os nomes dos personagens tivessem aparição relativamente destacada. A figura 2 apresenta o grafo deste artigo, produzido através da mineração do texto executada pelo programa Sobek. O prezado leitor pode comparar sua própria leitura com a mineração feita algoritmicamente pelo programa, que resultou no destaque dos termos e das setas indicando relação, conforme apresentado na figura 2.

Figura 2 – Grafo de mineração do texto do artigo “Humanidades Digitais e a simbiose entre humano e máquina – Algumas reflexões comparativas entre a interpretação e a mineração de textos”



Fonte: Elaborado pelos autores utilizando o programa Sobek (2019)

Poder-se-ia fazer a objeção de que a complexidade do Sobek é muito baixa, a ponto de permitir uma discussão mais detalhada da relação simbiótica entre leitura e mineração de textos. Mas as vantagens de apresentar o Sobek como exemplo são: Primeiro, que este programa vem sendo empregado em diversas iniciativas de aprendizagem da leitura e interpretação de textos, em processos educacionais que se caracterizam pela relação com a tecnologia digital. Uma dessas iniciativas está descrita no artigo “Estratégias pedagógicas de apoio à leitura, à escrita e ao acompanhamento baseadas na tecnologia de mineração de texto” de Macedo *et al.* (2011), citado acima. A segunda razão é que, para fins de sua própria experiência, qualquer leitor pode manusear o programa Sobek, e testar sua aplicação na mineração de textos, já que este não requer conhecimento de linguagens artificiais de programação (vide <http://sobek.ufrgs.br/>).

Outros programas de mineração de textos que se pode mencionar aqui são o Iramuteq (<http://www.iramuteq.org/>) e o NLTK - *Natural Language Tool Kit* (<https://www.nltk.org/>). Ambos são usados para fins de pesquisa científica, devido a seu alto desempenho. Contudo, exigem conhecimentos mais avançados de computação para serem manuseados. Não importa se a inteligência depositada nestes programas seja oriunda da inteligência de sucessivas etapas de projeto, por parte dos engenheiros que os projetaram. Para os fins a que se destinam, a saber, a análise e descoberta de informações não triviais em textos escritos, os programas de mineração de textos são casos daquilo que se chama ‘inteligência artificial’. Enquanto ferramenta de descoberta de ordem, padrões e regularidades, a mineração de textos é um

instrumento de análise que melhora a interpretação, mesmo que jamais *substitua* a interpretação.

6 CONSIDERANDO LEITURA E MINERAÇÃO DE TEXTOS EM UM ‘RELACIONAMENTO SIMBIÓTICO’

O fenômeno da existência e aplicabilidade de programas computacionais que revelam informações significativas em textos provoca inevitavelmente uma comparação entre a mineração de textos, feita por computadores, e a leitura, feita pela mente humana. Além da possibilidade de empregar a mineração de textos como ferramenta, substituindo o demorado esforço de análise de grandes volumes de textos, é interessante comparar a mineração de textos com a leitura, em termos de suas operações sobre o objeto-texto.

Isso é revelador do que a mente faz com o texto quando está lendo, e por uma teoria da leitura baseada nesta comparação, é possível explicar este comportamento complexo que se denomina interpretação. Este é o ponto crítico da presente discussão, a interpretação direta pela leitura não necessita transformação ou preparação prévia dos “dados”. Ela se baseia numa comunicação direta pela via da língua escrita, que dispensa intermediários. AY Liu, em seu artigo “The Meaning of Digital Humanities” (2013), fala de “linhas de interpretação geradas pela observação de máquinas¹⁰” (AY, 2013, p. 414, tradução nossa). O artigo, como o título revela, está interessado em discutir a questão do significado no cenário das HD.

Na passagem, o autor está se referindo a um desenvolvimento no desempenho da leitura, que é obtido pelo emprego da mineração de textos. Diz ele que

[...] um computador deve ser capaz de ler textos algoritmicamente e descobrir agrupamentos e padrões de palavras sem agir sobre o conceito inicial de um intérprete procurando confirmar um tema particular¹¹ (AY, 2013, p. 414, tradução nossa).

Seu ponto de vista é ousado: para Ay Liu a tecnologia digital abre possibilidades de construção do significado dos textos escritos, possibilitando desenvolvimentos que não seriam possíveis sem a simbiose entre o leitor e a máquina. Note-se o uso da expressão “ler algoritmicamente”, para se referir à descoberta que a mineração de textos faz de padrões relevantes e informações na estrutura do texto. Esta forma de se referir à mineração de textos parece ser compatível, em grande medida, com a suposição de uma simbiose entre a mente e o computador digital.

¹⁰ “[...] *lines of inter-pretation generated by machine observation*” (AY, 2013, p. 414).

¹¹ “[...] *a computer should be able to read texts algorithmically and discover word cohorts or clusters leading to themes without acting on an initial concept from an interpreter looking to confirm a particular theme*” (AY, 2013, p. 414).

Aqui é o momento de retomar esta noção de simbiose entre a mente humana e o computador, apresentada no início da discussão, e aplicá-la adequadamente ao caso da relação entre a leitura e a mineração de textos. O *Oxford English Dictionary* (OED) define o termo como: “Interação entre dois organismos diferentes vivendo em associação física próxima, tipicamente para a vantagem de ambos¹²” (OED). Essa noção específica da biologia é estendida a um sentido mais amplo e geral: “Um relacionamento mutuamente benéfico entre diferentes pessoas ou grupos¹³” (OED). Uma simbiose é uma relação entre dois organismos de espécies diferentes que, no curso da evolução, encontraram mútuo benefício em compartilhar os recursos do ambiente. Simbiontes são organismos que empregam uma forma de vida relacional, compartilhada, como estratégia adaptativa em longo prazo.

Não é sem razão que as ciências cognitivas, esta área interdisciplinar de estudo que se volta à compreensão da mente humana, conduzem suas pesquisas em proximidade com partes importantes da ciência da computação e da inteligência artificial. Howard Gardner, em sua obra *A nova ciência da mente* (1996), um clássico deste campo, menciona como um dos pilares das ciências cognitivas

[...] a crença de que o computador eletrônico é essencial para qualquer compreensão da mente humana. Os computadores não apenas são indispensáveis para a realização de estudos de vários tipos, mas, de forma mais crucial, o computador também é útil como o modelo mais viável de como a mente humana funciona (GARDNER, 1996, p. 20).

140

Esta característica das ciências cognitivas tem a ver com a revolução na tecnologia digital. Esta tecnologia é que permitiu o planejamento dos computadores e a possibilidade de utilizá-los para analisar dados, como ocorre com os textos em linguagem natural. Claro está que uma mente é algo radicalmente diferente de um computador. Mas as operações de ambos têm uma enorme semelhança que é causal, e não casual. Mas são as diferenças entre a mente e o computador, e o desempenho excelente de cada uma destas entidades em seu domínio específico, que justificam a alegação de uma relação simbiótica com propósitos coordenados.

Asimov formula a sua ideia de relacionamento simbiótico a partir da constatação de que

[...] o cérebro humano, composto de ácido nucléico e proteína em meio aquoso, resulta de três bilhões e meio de anos de evolução biológica, baseada em efeitos fortuitos da mutação, seleção natural e outras influências, e estimulada pela necessidade de sobrevivência (ASIMOV, 2005, p. 14).

A par disso, temos o computador digital que,

[...] composto por interruptores eletrônicos e corrente elétrica em meio metálico, resulta de quarenta anos de aperfeiçoamento da criação humana, baseada em

¹² “Interaction between two different organisms living in close physical association, typically to the advantage of both” (OED).

¹³ “A mutually beneficial relationship between different people or groups” (OED).

meticulosa previsão e na engenhosidade do próprio homem, e estimulada pela necessidade de servir a seus usuários” (ASIMOV, 2005, p. 14).

A proposta que se faz na presente reflexão é de que a mente humana e os computadores digitais sejam vistos como entidades simbióticas, considerando o vasto pano de fundo de um processo de evolução.

Trata-se aqui da evolução das próprias entidades, mas também do ambiente em que elas se encontram e no qual a simbiose é uma estratégia adaptativa. Em um trabalho intitulado “Fundamentos filosóficos para uma teoria evolutiva da informação e da cultura” (MATOS, 2016), esta ideia evolutiva e ambiental da atual sociedade da informação foi apresentada em maiores detalhes. No caso presente, este ambiente é a cultura, constituída de narrativas, conhecimentos, valores e mensagens. Mais especialmente, é a cultura considerada como o conjunto dos artefatos de todos os tipos, e também como uma espécie de herança, ou legado, cuja apropriação, transformação e transmissão constituem a vida cultural.

Norbert Wiener (2016), discute esta relação entre humano e máquina, na aurora da era digital, em sua obra *Cibernética* (publicada inicialmente em 1947). Na época, a situação do computador em relação à mente humana já dava margem a comparações, e o futuro da computação parecia ao autor, estar ligado a uma concepção da mente e da inteligência. Dizia ele que: “De há muito é claro para mim que o moderno computador ultrarrápido era em princípio um sistema nervoso central ideal para um aparelho de controle automático” (WIENER, 2016, p. 50). O computador digital logo se apresentou como uma metáfora do cérebro humano, e sua distinção essencial é o alto desempenho em operações rotineiras ou automáticas de processamento de informação.

Wiener (2016), além disso, antecipou uma futura convivência do computador com o humano, em que a divisão de tarefas implicasse no bem-estar das pessoas. Segundo suas palavras: “Pode ser ótimo para a humanidade que a máquina desobrigue do trabalho servil e desagradável, mas também pode não o ser” (WIENER, 2016, p. 50). Este resultado “ótimo para a humanidade”, conforme o autor exige o cultivo de princípios e valores humanísticos, a partir dos quais as decisões e a conduta individual e social possam se estruturar na direção de uma simbiose bem sucedida. “A resposta, sem dúvida, é ter uma sociedade baseada em outros valores humanos que não os da compra ou da venda” (WIENER, 2016, p. 51). Aqui se inaugura a questão moral do emprego da tecnologia informacional, pois a relação entre máquina e humano pode muito bem resultar em conflito, ou negligência - no caso das possibilidades de evolução tanto da tecnologia quanto da cultura humana não receberem a devida atenção nos estudos e iniciativas relevantes.

Autores como Dennett (1998), Floridi (2011) e Gleick (2013), apresentam em suas reflexões uma caracterização da sociedade digital como um meio ambiente. Dennett afirma: “Somos a única espécie que tem um meio extra de preservação e comunicação de projeto: a cultura” (DENNETT, 1998, p. 352). A cultura humana - sobretudo com o desenvolvimento do ciberespaço - é constitutiva de um tipo novo e desafiador de ambiente, em que a espécie humana procura se adaptar. Esta cultura digital é um ambiente em que a busca da adaptação pelos seus habitantes leva a relações interativas e a um processo complexo de evolução, nos termos do que se pode chamar uma *infosfera*.

No livro *A mente pós-evolutiva* (2010), João de Fernandes Teixeira considera que a evolução humana sofre os efeitos da civilização – e recentemente da tecnologia digital. A computação é examinada por ele à luz desta concepção num discurso que o autor denomina de “filosofia da mente no universo de silício” (TEIXEIRA, 2010). Suas conclusões sobre o futuro desenvolvimento da civilização apontam para o inevitável estreitamento da relação dos humanos com os computadores, bem como os avanços da inteligência artificial. Para Teixeira:

O mundo pós-evolutivo não será nem melhor nem pior que o atual. O desenvolvimento da tecnologia parece ser nossa única saída para sobreviver à involução. A parabióse, ou seja, nossa mistura com as máquinas, não é um aperfeiçoamento artificial da espécie humana, mas sim uma de suas últimas tentativas de sobreviver (2010, p. 154).

142

Esta concepção do efeito da relação com os computadores, sob um pano de fundo de desenvolvimento e evolução – biológica e cultural – tem ares de família com a proposta de uma integração simbiótica entre mente e máquina, na atividade tão fundamental e necessária da leitura. Poderia ser dito aqui, ao fim do exame de todas essas evidências e ideias, que as HD são muito mais do que um elenco de ferramentas e procedimentos para recuperação, análise e disseminação da cultura humanística. As HD são parte de uma estratégia mais ampla da própria sobrevivência e perspectiva de crescimento dessa cultura, amparada pelos recursos da tecnologia digital.

7 CONSIDERAÇÕES FINAIS

A vida cultural está migrando velozmente para uma sociedade da informação, ou infosfera, que se caracteriza pelo fluxo intenso e constante de grandes volumes de dados, especialmente pela via das redes digitais. Por sua vez, as HD podem ser consideradas como o campo de estudos e iniciativas envolvendo a promoção dos valores desta vida cultural na infosfera, ou seja, no ambiente digital. Admitindo a diferença essencial entre as funções mentais

e as funções das máquinas, a cooperação simbiótica entre mente e máquina favorece o desempenho de ambas. A saída simbiótica, considerando a relação entre a leitura e a mineração de textos, e sua ampla difusão nas práticas de pesquisa e projetos das HD, contribui para a adaptação dos seres humanos ao ambiente mutável de sua própria cultura. Dada a vastidão insondável desta cultura, a simbiose é um recurso adaptativo que, hoje em dia, parece cada vez mais irreversível. O caminho evolutivo dos próprios computadores, e a modelagem da inteligência artificial que estes incorporam, estão determinados por esta relação.

Na medida em que a discussão das questões aqui propostas desperte o interesse da comunidade de estudiosos das HD, abre-se a possibilidade de que a relação entre a leitura e a mineração de textos seja alvo de investigação científica mais específica. O presente trabalho terá cumprido sua finalidade se puder incitar a indagação dos pesquisadores na direção desta forma de conceber a situação da leitura e da mineração de textos frente aos princípios e valores teóricos das HD.

Para além dos limites desta discussão, permeada de ligações com a literatura e a filosofia, permanece aberta a possibilidade de pesquisas mais pontuais. Um levantamento exaustivo de projetos de HD em que a mineração de textos é empregada em correlação com os hábitos e procedimentos de leitura, com uma análise crítica dos resultados obtidos em tais projetos, por exemplo, seria uma contribuição relevante para esta discussão.

Por último, deixa-se o prezado leitor com uma provocação, motivada por todo o percurso teórico e argumentativo apresentado acima: A relação simbiótica entre a leitura e a mineração de textos é um caso específico da relação simbiótica mais geral entre a mente humana e o computador digital. Esta relação, considerando todo o discurso desenvolvido até aqui, poderia ser considerada como o método geral das HD. Em outras palavras, a prática das HD, a julgar pelas afirmações de seus intérpretes, é o desenvolvimento de diversas iniciativas particulares, cuja forma geral é a de uma simbiose entre o ser humano e o computador digital. Uma simbiose cujos benefícios e os riscos sequer começam a ser observados com nitidez, na rápida evolução cultural a que todos estamos submetidos.

REFERENCIAS

ARANHA, Christian; PASSOS, Emmanuel. A tecnologia de mineração de textos. **RESI: Revista eletrônica de sistemas de informação**, Curitiba, v. 5, n. 2. 2006. Disponível em: <http://www.periodicosibepes.org.br/index.php/reinfo/article/view/171>. Acesso em 25 de janeiro de 2019.

ASIMOV, Isaac. Os robôs, os computadores e o medo. *In*: WARRICK, Patricia S.; GREENBERG, Martin H. (orgs.). **Histórias de Robôs**. Porto Alegre: L&PM, 2005. v. 2.

CUARTAS-RESTREPO, J. M. Humanidades digitais, dejarlas ser. **Revista Colombiana de Educación**, Bogotá, n. 72, p. 65-78. 2017.

DENNETT, Daniel. **A Perigosa ideia de Darwin**. São Paulo: Rocco, 1998.

DESCARTES, René. **Discurso do Método**. São Paulo: Nova Cultural, 1984.

FLORIDI, Luciano. **The Philosophy of Information**. London: Routledge, 2011.

GLEICK, James. **A informação: uma teoria, uma história, uma enxurrada**. São Paulo: Cia das Letras, 2013.

LEYVA, Elsa M. Ramírez. ¿Qué es leer? ¿Qué es La lectura? **Investigación Bibliotecológica**, México. v. 23, n. 47, enero/abril. 2009.

LOBATO, Monteiro. **O Presidente Negro**. São Paulo: Globo, 2009.

MACEDO, Alexandra Lorandi *et al.* Estratégias pedagógicas de apoio à leitura, à escrita e ao acompanhamento baseadas na tecnologia de mineração de texto. *In*: SIMPÓSIO BRASILEIRO DE INFORMÁTICA NA EDUCAÇÃO – SBIE, 22., 2011, Aracaju, Anais do XXII SBIE - XVII WIE, Aracaju: SBIE, 2011. p. 1017-1026.

MANOVICH, Lev. The Science of Culture? Social computing, Digital Humanities and Cultural Analytics. **Journal of cultural Analytics**, May 23. 2016.

MATOS, José Claudio. A interpretação de textos e a formação da pessoa reflexiva – sobre a concepção deweyana da leitura. **Educação e Filosofia**, Uberlândia, v. 27, n. 54, p. 579-596, jul./dez. 2013.

MATOS, José Claudio. Fundamentos filosóficos de uma teoria evolutiva para a informação e a cultura. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., 2016, Salvador. **Anais [...]**. v. 1. p. 1-20. João Pessoa: UFPB, 2016.

MORAIS, Edison Andrade Martins; AMBRÓSIO, Ana Paula L. **Mineração de textos**. Goiás: UFG, 2007. (Relatório Técnico. Instituto de Informática). Disponível em:

http://www.inf.ufg.br/sites/default/files/uploads/relatorios-tecnicos/RT-INF_005-07.pdf. Acesso em 25 de janeiro de 2019.

MORAIS, José. **A Arte de ler**. Lisboa: Edições 70. 2002.

PRIANI, E. El texto digital y ladisyuntiva de las humanidades digitales. **Palabra Clave**, Buenos Aires, v. 18, n. 4, p. 1215-1234. 2015.

SÁNCHEZ, Adriana Álvarez. La diversidad Cultural em las HD. *In*: PRIANI SAISO, Ernesto *et al.* **Humanidades Digitales: Lengua, texto, Patrimonio**. Mexico: Bonilla y Red de Humanidades Digitales, 2018. p. 183-208.

SYMBIOSIS. *In*: OXFORD English Dictionary. Versão online. [S. l.]: Disponível em: <https://en.oxforddictionaries.com/definition/symbiosis>. Acesso em 30 de janeiro de 2019.

TĂŪT, Val Codrin. Beyond the tale of the ‘two cultures’: filling the gap between algorithms and interpretation. *In*: INTERNATIONAL SCIENTIFIC CONFERENCE E LEARNING AND SOFTWARE FOR EDUCATION, 13., 2017, Bucharest, **Proceedings** [...]. [S.l.]; Bucharest, 2017.

TEIXEIRA, João de Fernandes. **A mente pós-evolutiva: filosofia da mente em um universo de silício**. Petrópolis: Vozes, 2010.

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL. **Sobek quick reference guide**. Grupo de Pesquisa Gtech.edu. UFRGS, 2017.

WIENER, Norbert. **Cibernética: o controle no animal e na máquina**. São Paulo: Cia das Letras. 2016.