

Estudo exploratório sobre “Digital Humanities” na base de dados Scopus com aplicação do software IRaMuTeQ

Exploratory study on “Digital Humanities” in the Scopus database with IRaMuTeQ software application

Priscila Ramos Carvalho¹, Marcos Gonçalves Ramos², Ricardo Medeiros Pimenta³

Resumo

O presente estudo é o relato da pesquisa empírica realizada na base de dados Scopus visando identificar termos que refletem a heurística das Humanidades Digitais por meio da análise dos resumos de artigos recuperados de 1990 a 2019, através do programa IRaMuTeQ. O resultado apontou a transdisciplinaridade das Humanidades Digitais com a Ciência da Informação e a História evidenciada pelos termos “library”, “librarian”, “technology”, “datum” e “history”.

Palavras-chave: Humanidades Digitais, Scopus, IRaMuTeQ

Abstract

The present study is the report of the empirical research carried out in the Scopus database aiming to identify terms that reflect the heuristic of the Digital Humanities through the analysis of the abstracts of articles retrieved from 1990 to 2019, through the IRaMuTeQ program. The result pointed to the transdisciplinarity of the Digital Humanities with Information Science and History evidenced by the terms “library”, “librarian”, “technology”, “datum” and “history”.

Keywords: Digital Humanities, Scopus, IRaMuTeQ

1 Introdução

As Humanidades Digitais (HD) referem-se ao encontro transdisciplinar entre as áreas das Ciências Humanas e Sociais, às Artes e às Letras, no contexto da virada digital desde o final do século XX, potencializada pelo advento da internet e digitalização de todas as formas de expressão da cultura e do conhecimento humanos.

¹ Doutoranda no Programa de Ciência da Informação no IBICT/ECO-UFRJ.

² Doutoranda no Programa de Ciência da Informação no IBICT/ECO-UFRJ.

³ Pesquisador do IBICT e Professor do Programa de Ciência da Informação no IBICT/ECO-UFRJ.

De acordo com Palleta (2018), as HDs incluem o uso sistemático de recursos, ferramentas e métodos digitais voltados para a reflexão sobre sua aplicação de tecnologias digitais que compreendem um conjunto de valores e de atitudes provenientes de investigações e iniciativas práticas na academia.

O Manifesto das Humanidades Digitais⁴ destaca o caráter múltiplo e interativo das comunidades que participam da construção de 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 podem convergir para a construção de um campo de conhecimento.

Diante da multiplicidade de atores e conhecimentos, seria possível identificar um núcleo semântico composto de termos que refletissem a heurística das Humanidades Digitais? A partir desta questão de pesquisa, foi realizado um estudo empírico na base de dados Scopus visando a coleta de artigos por meio do termo “Digital Humanities” nos últimos 30 anos, a fim de realizar a análise dos resumos recuperados através do programa de análise textual IRaMuTeQ.

2 Metodologia

A pesquisa empírica-descritiva tomou como base a perspectiva das metrias da informação, utilizando a Cientometria como metodologia que aplica análises quantitativas sobre a produção da ciência enquanto uma disciplina ou atividade econômica, que envolve entre outras atividades científicas, a publicação de artigos. (MACIAS CHAPULA, 1998, p.134).

O estudo teve como objetivo identificar os termos mais frequentes relacionados às “Digital Humanities” nos resumos de artigos recuperados na base de dados Scopus, no período de 1990 a 2019 (30 anos). A coleta de dados foi realizada no dia 30 de janeiro de 2020, tendo como resultado 2366 documentos, porém após o tratamento dos dados verificou-se que 208 não possuíam resumos.

Tabela 1 – Resultado da Expressão de Busca na Base de Dados Scopus (1990-2019)

Termo	Expressão de Busca	Resultado
Digital Humanities	TITLE-ABS-KEY ("digital humanities") AND PUBYEAR > 1989 AND PUBYEAR < 2020	2366

Fonte: elaborado pelos autores com base na Scopus

⁴ Manifesto for the Digital Humanities. Disponível em: < <https://humanidadesdigitais.org/manifesto-das-humanidades-digitais/> > Acesso em: 10 dez. 2020.

Em seguida, o total de 2158 resumos foi consolidado em um *corpus textual* para análise no software de código aberto e gratuito IRaMuTeQ (*Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires*).

O programa escolhido foi criado por Pierre Ratinaud, no Laboratório de Estudos e Pesquisa Aplicada em Ciências Sociais em Toulouse, nas linguagens estatísticas R e Python, possibilitando diferentes análises textuais como, por exemplo, as lexicais clássicas que identificam a quantidade total, a frequência média e o número de palavras com frequência igual a um. Além disso, a ferramenta permite a pesquisa de vocabulário e faz lematização, assim como cria um dicionário de formas reduzidas e identifica formas ativas e suplementares (CAMARGO; JUSTO, 2013).

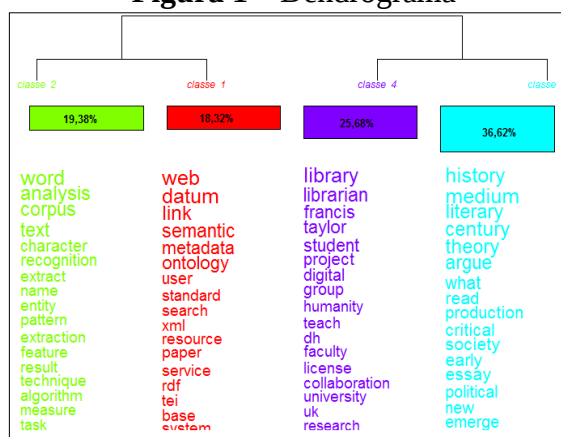
A lematização envolve a busca e a relação de palavras, agrupando-as de acordo com alguns critérios como o tratamento de plurais para singulares, as formas de gênero e as diferentes flexões, reduzindo as palavras com base no radical. (SALEM, 1986).

3 Resultado da Análise de Dados da Scopus

A primeira análise foi estatística simples do *corpus textual* que apresentou: 8.602 segmentos de texto, 352.673 ocorrências de palavras, 13.176 palavras ativas (substantivo, adjetivo e advérbio), 897 palavras suplementares (artigos, pronomes, outros), 6.219 palavras com uma única ocorrência e 4 classes.

A segunda análise foi a classificação hierárquica descendente que gerou o dendrograma retratado na figura 1 composto por: 1) classe 1 vermelho (18,32% em relação ao total) com destaque para os seguintes termos: “web”, “datum”, “link”, “semantic”; 2) classe 2 verde (19,38%) os termos mais relevantes: “word”, “analysis”, “corpus”; 3) classe 3 azul (36,62%) com termos mais salientes: “history”, “medium”, “literacy”; 4) classe 4 roxa (25,68%) com realce para os termos: “library”, “librarian”, “francis”.

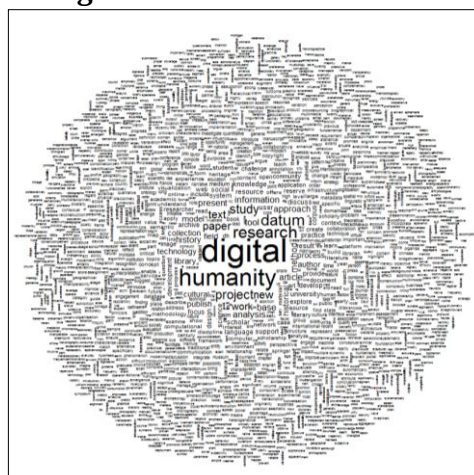
Figura 1 – Dendrograma



Fonte: elaborado pelos autores através do IRaMuTeQ

A terceira análise foi sobre a frequência dos termos representada pela nuvem de palavras na figura 2, a qual mostrou o termo “digital” como central. Alguns termos merecem também destaque, tais como: “humanity”, “research”, “datum”, “study”.

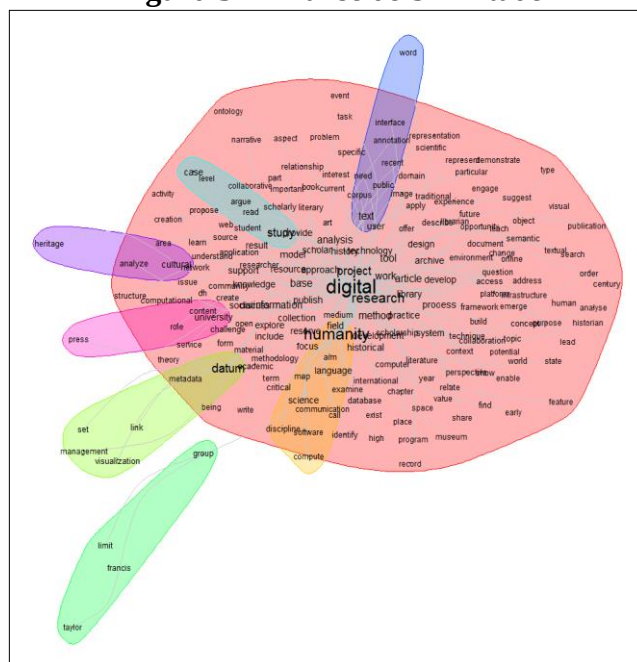
Figura 2 – Nuvem de Palavras



Fonte: elaborado pelos autores através do IRaMuTeQ

A quarta foi a análise de similitude baseada na teoria dos grafos que exibiu um cluster principal simbolizado pelo termo “digital”, além de ser um cluster formado por mais sete subcluster identificados pelos seguintes termos: “humanity”, “study”, “text”, “university”, “cultural”, “datum”, “web”. Convém acrescentar que o termo “digital” possui atração semântica com os termos: “research”, “project”, “technology”, “library”, “history”, “method”.

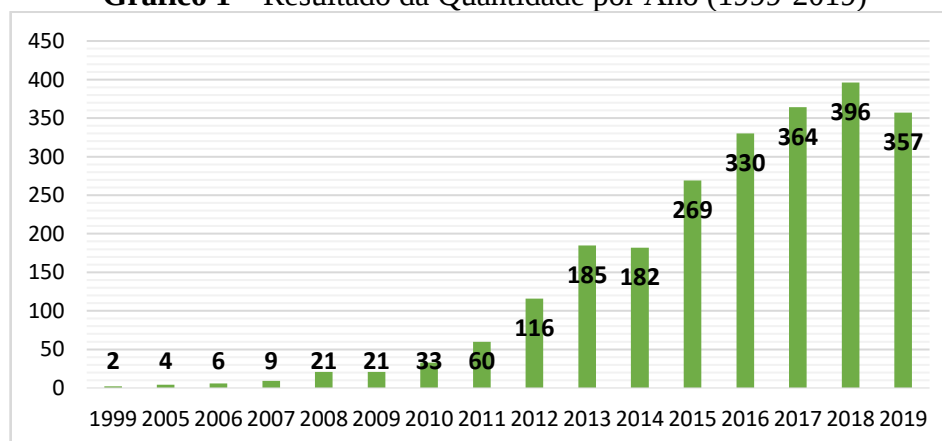
Figura 3 – Análise de Similitude



Fonte: elaborado pelos autores através do IRaMuTeQ

A fim de complementar a percepção da recuperação da informação na Scopus foi analisado o volume da produção por ano, retratado no gráfico 1, assim como os países e áreas de assunto (*subject area*).

Gráfico 1 – Resultado da Quantidade por Ano (1999-2019)



Fonte: elaborado pelos autores

A primeira indexação de “Digital Humanities” foi em 1999 com 2 documentos e a partir de 2012 houve crescimento de publicações significativo de 93,33%, saindo de 60 em 2011 para 116 em 2012.

Em relação aos países, os Estados Unidos se destacaram em primeiro lugar com 751 documentos, em segundo a Grã-Bretanha com 274, e em terceiro a Alemanha com 238. No que se refere às áreas de assunto sinalizadas na Scopus, em primeiro lugar foi “Social Sciences”

com 1.191 documentos, em segundo “Computer Sciences” com 1.101, em terceiro “Arts and Humanities” com 1.022.

4 Considerações Finais

O resultado preliminar mostrou a transdisciplinaridade das Humanidades Digitais com a Ciência da Informação e a História ao perceber termos como “*library*”, “*librarian*”, “*technology*”, “*datum*”, “*history*”. De acordo com Berdan (2013), às Humanidades Digitais representam um trabalho realizado na interseção de tecnologia e humanidades.

A intenção é ampliar o estudo para um comparativo com outras bases de dados como, por exemplo, a Library Information Science Abstracts – LISA, a Web of Science e a Dimensions.ai de maneira a perceber a ligação das Humanidades Digitais com diferentes áreas do conhecimento e metodologias digitais.

Referências

BERDAN, J. The Emerging Field of Digital Humanities: An Interview with Johanna Drucker. **InterActions**: UCLA Journal of Education and Information Studies. v.9 n.2, 2013.

CAMARGO, B. V.; JUSTO, A. M. IRAMUTEQ: um software gratuito para análise de dados textuais. **Associação Brasileira de Psicologia**. Temas em Psicologia, Ribeirão Preto, v. 21, n. 2, p.513-518, 2013.

PALETTA, F.C. Ciência da informação e humanidades digitais – uma reflexão. **Encontro Nacional de Pesquisa em Ciência da Informação**, In. XIX ENANCIB, 2018. Disponível em: <<http://hdl.handle.net/20.500.11959/brapci/103662>>. Acesso em: 10 dez. 2020