

Ricardo César Gonçalves Sant'Ana

Moisés Lima Dutra

Guilherme Ataíde Dias

Organizadores

WIDaT 2018

II WORKSHOP DE INFORMAÇÃO,
DADOS E TECNOLOGIA

ANAIS
WIDaT 2018

Organização do WIDaT 2018

- **Organização Geral:**

Guilherme Ataíde Dias (PPGCI-UFPB) - Coordenador geral do evento
Moisés Lima Dutra (PPGCIN-UFSC) - Vice-coordenador

- **Coordenador da Comissão Científica:**

Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)

- **Comissão científica**

Adilson Luiz Pinto (PPGCIN-UFSC)
Ana Alice Baptista (Universidade do Minho, Portugal)
Ana Carolina Simionato (PPGCI-UFSCar)
Angela Maria Grossi de Carvalho (PPGCI-UNESP)
Bernardina Maria Juvenal Freire de Oliveira (PPGCI-UFPB)
Cristian Berrío-Zapata (PPGCI-UFPA)
Dalton Lopes Martins (FCI-UnB)
Denysson Axel Ribeiro Mota (PPGB-UFCA)
Douglas Dyllon Jeronimo de Macedo (PPGCIN-UFSC)
Ed Porto Bezerra (PPGI-UFPB)
Edgar Bisset Alvarez (PPGCIN-UFSC)
Edna Gusmão de Goés Brennand (MPGOA-UFPB)
Edna Gomes Pinheiro (DCI-UFPB)
Elaine Parra Affonso (FATEC-SP)
Elvis Fusco (UNIVEM-Marília)
Enrique Muriel Torrado (PPGCIN-UFSC)
Evandro de Barros Costa (IC-UFAL)
Fábio Paraguaçu (IC-UFAL)
Fernando de Assis Rodrigues (PPGCI-UNESP)
Gustavo Medeiros de Araújo (PPGCIN-UFSC)
Henry Pôncio Cruz de Oliveira (PPGCI-UFPB)
Joana Coeli Ribeiro Garcia (PPGCI-UFPB)
José Eduardo Santarém Segundo (USP-FFCLRP)
Leonardo Castro Botega (UNIVEM-Marília)
Luana Farias Sales Marques (PPGCI-IBICT-UFRJ)
Marckson Roberto Ferreira de Sousa (PPGCI-UFPB)
Luís Fernando Sayão (CNEN)
Marcelo Morandini (EACH-USP)
Márcio Matias (PPGCIN-UFSC)
Marcos Mucheroni (CBD-USP)
Marynice de Medeiros Matos Autran (PPGCI-UFPB)

Maurício Barcellos Almeida (PPGGOC-UFMG)
Moisés Lima Dutra (PPGCIN-UFSC)
Plácida Leopoldina V. da Costa Santos (PPGCI-UNESP)
Pedro Luiz Pizzigatti Corrêa (POLI-USP)
Renata Baracho (PPGGOC-UFMG)
Ricardo César Gonçalves Sant'Ana (PPGCI-UNESP)
Robson Rodrigues Lemos (UFSC-Araranguá)
Rogério Ramalho (PPGCI-UFSCar)
Ryan Ribeiro de Azevedo (UFRPE-UAG)
Sandra de Albuquerque Siebra (PPGCI-UFPE)
Sandro Rautenberg (DECOMP-UNICENTRO)
Silvana Aparecida Borsetti G. Vidotti (PPGCI-UNESP)
Virginia Bentes Pinto (PPGCI-UFC)
Wagner Junqueira de Araújo (PPGCI-UFPB)
Zaira Regina Zafalon (PPGCI-UFSCar)

- **Coordenador do Cerimonial:**

André Luiz Dias de França (PPGCI-UFPB)

- **Coordenador da Equipe Técnica Local:**

Laerte Pereira da Silva Júnior (CCHLA-UFPB)

- **Equipe Técnica Local:**

Adriana Alves Rodrigues (PPGCI-UFPB)
Antonio Felipe dos Santos (MPGOA-UFPB)
Débora Gomes de Araújo (PPGCI-UFPB)
Pedro Augusto de Lima Barroso (PPGCI-UFPB)
Pollianna Marys de Souza e Silva (PPGCI-UFPB)
Renata Lemos dos Anjos (PPGCI-UFPB)

HUMANIDADES DIGITAIS: visualização da produção científica

*DIGITAL HUMANITIES:
visualization of scientific production*

Mirelys Puerta Díaz¹, Edgar Bisset Alvarez², Silvana Aparecida Borsetti Gregorio Vidotti³

(1) Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, Campus Marília), mireblue18@gmail.com, mirelys.puerta@unesp.br

(2) Universidade Federal de Santa Catarina, Florianópolis, edgar.bisset@ufsc.br

(3) Universidade Estadual Paulista “Júlio de Mesquita Filho” (UNESP, Campus Marília), silvana.vidotti@unesp.br

Resumo:

O presente artigo aborda as Humanidades Digitais através da produção científica publicada nas bases de dados científicas nacionais e internacionais. No estudo, utilizou-se o método da bibliometria aplicada para identificar a produção científica indexada sobre o objeto pesquisado. A pesquisa é descritiva e quali-quantitativa. Utilizou-se como fonte, a Base de dados de Periódicos em Ciência da Informação e documentos recuperados com o Google Acadêmico através do software Publish or Perish, visando verificar a produção teórica das HD e as métricas de interesse. Aplicou-se para a visualização das métricas, o aplicativo *RAW Graphs*, baseado na web, trata-se de um framework de código aberto utilizado para a visualização de dados complexos. Com isso, foi possível constatar que no Brasil, os estudos sobre humanidades digitais na Ciência da Informação foram publicados somente a partir do ano 2014, enquanto no contexto internacional, as publicações já vinham sendo publicadas desde aproximadamente o ano 2000.

Palavras-chave:

Humanidades digitais. Visualização de dados. Análise métrica da produção científica.

Abstract:

This article addresses the Digital Humanities through the scientific production published in national and international scientific databases. In the study, the applied bibliometric method was used to identify the indexed scientific production on the researched object. The research is descriptive and qualitative-quantitative. It was used as source, the Database of Periodicals in Information Science and documents retrieved with Google Academic through the software Publish or Perish, to verify the theoretical production of the HD and the metrics of interest. It was applied to the visualization of the metrics, the web-based *RAW Graphs* application, it is an open-source framework used for the visualization of complex data. With this, it was possible to verify that in Brazil, studies on digital humanities in Information Science were only published as from 2014, while in the international context, publications had been published since about 2000.

Keywords:

Digital Humanities. Data visualization. Scientific production metric analysis.

I INTRODUÇÃO

Nos últimos anos tem se observado um crescente avanço das pesquisas que envolvem as Humanidades Digitais (HD), o que possibilitou defini-las como campo disciplinar, perfil formativo ou como saber ligado ao movimento do livre acesso. E, esta produção científica também se ampliou a nível internacional, e, de tal maneira, que na opinião de Rodríguez-Yunta (2012), atualmente as HD podem ser consideradas como uma *topic tag* de significado ambíguo, cujo exponencial crescimento deve motivar a reflexão dos profissionais da Ciências da Informação.

Deste modo, esta área de estudo tem se convertido em verdadeira fonte para o desenvolvimento de projetos apontados como ponto de encontro entre as Ciências da Computação, o Ambiente digital e as Ciências Humanas e Sociais, neste sentido, Ortega (2013, p.18):

“[...] as Humanidades Digitais podem ser definidas como o espaço de convergência entre as ciências da computação, meio digital e as disciplinas humanísticas na busca de novos modelos interpretativos e novos paradigmas de conhecimento em concordância com as transformações operadas no âmbito da sociedade digital.”

Entretanto, apesar da crescente popularidade deste campo, não existe consenso no que diz respeito às suas bases teórico-epistemológicas, muito menos, quanto à construção de uma definição que resuma a essência do que são as HD, como aliás, fica evidenciado, no levantamento feito por Jason Heppler¹ em 2015, que reuniu cerca de 800 definições diferentes sobre o que são as HDs.

2 OBJETIVOS

O presente trabalho tem como objetivo geral apresentar o comportamento da produção científica das Humanidades Digitais², a partir da análise dos dados coletados nas bases de dados Google Acadêmico e a Base de dados de Periódicos em Ciência da Informação (BRAPCI).

3 PROCEDIMENTOS METODOLÓGICOS

A presente pesquisa descritiva combina as abordagens qualitativa e quantitativa e aplica-se a análise métrica da produção científica para alcançar seu objetivo.

A coleta de dados foi realizada sem limitações na data de publicação na BRAPCI, e através do buscador Google Acadêmico utilizando o software *Publish or Perish* (PoP), onde limitou-se a busca aos anos 2000-2017. O uso de PoP se justifica porque a interface nativa de Google Acadêmico não é muito adequada para análises bibliométricas.

Por meio da coleta na BRAPCI, se pretende identificar, além da presença das Humanidades Digitais nas pesquisas da Ciência da Informação brasileira, o foco de suas abordagens. Com esse fim a busca foi realizada em língua portuguesa (“humanidades digitais”). O Método de pesquisa aplicado para a análise dos dados é o de análise bibliométrica da produção científica.

¹ https://github.com/hepplerj/whatisdigitalhumanities/blob/master/dayofquotes_full.csv

² Digital humanities, em sua versão em inglês.

A coleta, processamento e análise de dados em Google Acadêmico, foi apoiada pelo uso da ferramenta PoP, que recupera e faz análise de citações acadêmicas obtidas de fontes como Google Acadêmico. Da variedade de fontes de dados que usa PoP (por exemplo, *Crossref*, Google Acadêmico, *Web of Science y Scopus*) foi escolhida a do Google Acadêmico (GA) porque é considerada uma útil ferramenta para pesquisa de literatura científica de forma ampla, em muitas disciplinas e fontes acadêmicas (artigos revisados por pares, teses, livros, resumos e artigos de editoras acadêmicas, sociedades profissionais, repositórios de pré-impressão, repositórios de universidades e outras organizações acadêmicas). O PoP facilita o cálculo de várias métricas³ (HARZING, 2011), das quais, foram selecionadas para análise, as especificadas no Quadro 01:

Quadro 01

Métrica	Descrição
Anos de publicação	Primeiro e mais recente ano de publicação encontrado nos resultados atualmente selecionados.
Anos de citação	Número de anos, desde o primeiro ano encontrado nos resultados atualmente selecionados até o ano da consulta (geralmente o ano corrente).
Documentos	Número total de resultados atualmente selecionados.
Quantidade de citações recebidas	A soma das contagens de citações em todos os resultados atualmente selecionados.
Promedio de citações por año (citações/ano)	Número médio de citações por ano (ou seja, citações / anos de citação)
Promedio de citações por documento (Citações / documento)	A soma das contagens de citações em todos os artigos, dividida pelo número total de artigos. A média e o modo também são calculados e disponibilizados separadamente.
Citações / autor	Número médio ⁴ de citações por autor
Artigos / autor	Número médio ⁵ de artigos por autor.
Autores /documento	Número médio de autores ⁶ por artigo.
h-index ⁷	Índice H de Hirsch
índice g ⁸	Índice g de Egghe

Fonte: Research in International Management, 2017.

Os resultados obtidos com o programa ficam disponíveis na tela e também podem ser copiados para a área de transferência do Windows, para colar em outros aplicativos ou salvar em vários formatos de saída (*vector graphic (svg)*, *image (png)*, *data model (json)*) para futura referência ou análise posterior. Sendo que, a capacidade de exportar o código embutido em SVG e o modelo de

3 <https://harzing.com/resources/publish-or-perish/manual/using/query-results/metrics>

4 Calculado da seguinte forma: Para cada trabalho, sua contagem de citações é dividida pelo número de autores para esse artigo, para devolver a contagem de citações normalizada por autor para o artigo. As contagens de citações normalizadas são então somadas em todos os artigos para fornecer o número de citações por autor sobre o conjunto de resultados.

5 Calculado da seguinte forma: Para cada trabalho, a contagem de 1 / autor é calculada para fornecer a contagem de autor normalizada para o trabalho. A contagem normalizada dos autores é então resumida em todos os artigos para fornecer o número de artigos por autor.

6 Calculado como a soma do autor conta em todos os artigos, dividido pelo número total de artigos. A média e o modo também são calculados e disponibilizados separadamente.

7 Métrica que mede o impacto cumulativo da publicação de um pesquisador, observando a quantidade de citações que seu trabalho recebeu. O índice h foi proposto por J.E. Hirsch no ano 2005.

8 O índice g é um índice para quantificar a produtividade na ciência, com base no registro de publicação (uma métrica em nível de autor). O índice é calculado com base na distribuição de citações recebidas pelas publicações de um determinado pesquisador, de tal forma que dado um conjunto de artigos classificados em ordem decrescente do número de citações que eles receberam, o índice g é o maior número único de tal forma que o topo g artigos recebidos juntos pelo menos g 2 citações.

dados JSON é compatível com essa abordagem aberta, o que permite explorar novos usos para esses produtos, facilitando o surgimento de novos métodos de visualização dos dados inseridos.

Aplicou-se para a visualização dos dados resultantes da coleta com o PoP, o aplicativo baseado na web, *RAWGraphs*⁹; um framework de código aberto utilizado para a visualização de dados complexos, desenvolvido no Laboratório de Pesquisa *DensityDesign* (Politecnico di Milano), cria visualizações baseadas em vetores personalizados da biblioteca *d3.js*. O *RAWGraphs* funciona com valores separados por delimitador (ou seja, arquivos csv e tsv), bem como, com textos copiados e colados de outros aplicativos (por exemplo, *Microsoft Excel*, *TextWrangler*, *TextEdit*). E, com base no formato svg, as visualizações podem ser facilmente importadas e editadas com aplicativos de gráficos vetoriais (ou seja, *Adobe Illustrator*, *Inkscape*, entre outros) para refinamentos adicionais ou diretamente incorporadas em páginas da web. Sendo que, o aplicativo *Inkscape* foi utilizado para melhorar a aparência dos gráficos gerados pelo aplicativo web *RAWGraphs*.

RAWGraphs surge como uma das ferramentas mais favoráveis para usar tanto em termos de características operacionais quanto características funcionais para a visualização de dados (ATWOOD e REZNIK-ZELLEN, 2018, p18).

Neste estudo foram utilizadas as seguintes visualizações do *RAWGraphs*:

- Gráfico de Dispersão (*Scatter Plot*): facilita criar Círculos aninhados para representar hierarquias e comparar valores. Essa visualização é particularmente eficaz para mostrar a proporção entre elementos através de suas áreas e posição dentro de uma estrutura hierárquica.
- Círculo *Packing*: Círculos aninhados permitem representar hierarquias e comparar valores. Essa visualização é particularmente eficaz para mostrar a proporção entre elementos através de suas áreas e sua posição dentro de uma estrutura hierárquica.

Foram identificadas como limitações do presente estudo, erros ou omissões referentes ao uso do PoP no Google Acadêmico, as quais são mais frequentes do que no Scopus e no *Web of Science*. O uso do programa requer a limpeza da lista de publicações obtidas na busca, eis que, também coletou registros que não correspondem com a estratégia de busca. Além disso o Google Scholar limita seus resultados a 1000 registros, os quais são classificados por número de citações recebidas, portanto, os 1000 coletados pelo PoP são os resultados mais citados, sendo assim omitidos muitas publicações que ainda não foram citadas.

4 RESULTADOS

Verificou-se com o levantamento bibliográfico realizado na base de dados BRAPCI, que trouxe como resposta, apenas, um total de 9 documentos, dos quais, 5 foram publicados no ano de 2017, o raso aprofundamento da temática pela comunidade acadêmica no Brasil.

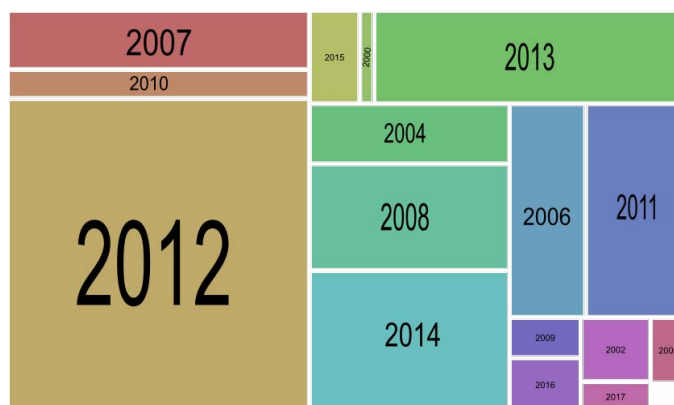
Sendo que, o principal foco das pesquisas desenvolvidas no cenário brasileiro abordam as humanidades digitais relacionadas à temáticas como, as tecnologias da informação e as comunicações, suas abordagens desde a Ciência da Informação, sua relação com a mediação, a web, a ciência aberta, entre outros que estão representados na nuvem de palavras da Figura 1:

⁹ O projeto, liderado e mantido pelo Laboratório de Pesquisas *DensityDesign* (Politecnico di Milano) foi lançado publicamente em 2013 e é considerado como uma das ferramentas mais importantes no campo da visualização de dados.

A Figura 3 demonstra que os autores que mais citações receberam (NK Hayles, NK Hayles, L Manovich, R Kitchin, P Juola, entre outros) encontram-se no centro da visualização, o que significa que quanto maior o número de citações, maior raio de cobertura dos autores que estão localizados no centro do gráfico.

O ano de maior quantidade de publicações sobre humanidades digitais foi o 2012 com um total de 166 trabalhos; seguido do ano 2013 com 122 publicações e 2014 com 103. Um fato interessante que nos últimos 3 anos diminuiu consideravelmente a produção científica de maior impacto sobre a temática segundo a fonte consultada em relação a anos anteriores. Do ano 2015 ao 2017 diminuiu a metade do total de publicações.

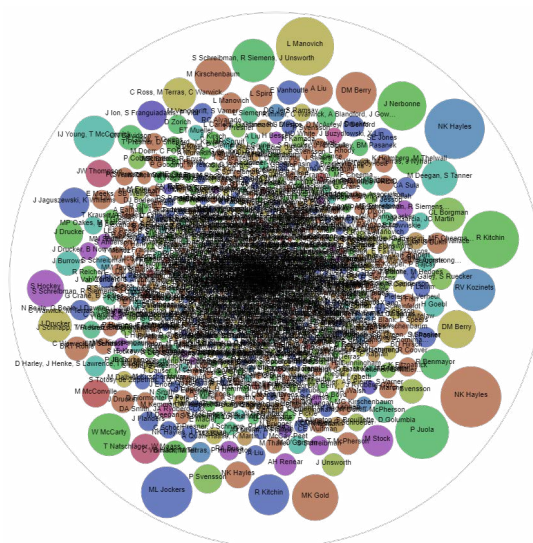
Figura 4: Visualização da métrica dos totales de produções científicas por ano.



Fonte: Elaboração própria por meio da ferramenta RAWGraphs.

A Figura 4 mostra a relação das dimensões ano de publicação e o total de citas recibidas por ano. Quanto maior o número de citações maior raio de cobertura dos anos que estão localizados no gráfico. Essa visualização das métricas mencionadas denota, com exceção do ano 2011 e 2008, que os três anos mais produtivos foram os mais citados.

Figura 5: Visualização das métricas do total de citações recebidas pelos autores e o GSR (Google Scholar Rank)



Fonte: Elaboração própria por meio da ferramenta RAWGraphs.

A visualização da Figura 5 denota que, os 5 trabalhos mais citados não aparecem nas 10 primeiras posições do ranking do Google Acadêmico, dos quais o 50% deles são publicações do ano 2012. Uma relação detalhada dos dados em relação aos autores mais citados está listada abaixo:

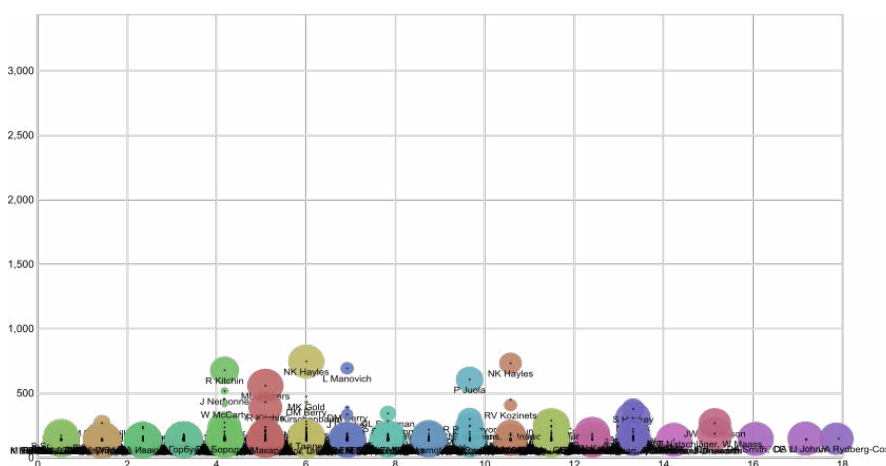
- NK Hayles (citações= 762; ano= 2012; GSRank =848)
- NK Hayles (citações= 746; ano= 2007; GSRank =315)
- L Manovich (citações= 700; ano= 2011; GSRank =95)
- R Kitchin (citações= 682; ano= 2014; GSRank =542)
- P Juola (citações= 598; ano= 2008; GSRank =485)
- ML Jockers(citações= 543; ano= 2013; GSRank =861)

Identificou-se que somente o 25 % dos 10 primeiros resultados do Google Acadêmico estão em acesso aberto e disponíveis para a comunidade científica.

Essa visualização das métricas mencionadas, denota, que não necessariamente os artigos mais citados pelos autores (exemplo. Cuddon, Hayles e Röhle) são colocados nos primeiros 10 resultados da pesquisa feita pelos usuários no Google Acadêmico. Outro fato interessante é que, os autores com mais citações são os do ano 2012, como foi identificado na Figura 3.

Também se verifica, que as publicações correspondentes a 2012 estão entre as mais citadas.

Figura 6 Visualização de métrica da idade das publicações e do total de citações recebidas.



Fonte: Elaboração própria por meio da ferramenta RAWGraphs.

O gráfico da Figura 6, mostra a correlação entre a dimensão da idade das publicações e o total de citações recebidas. O Eixo X= idade das publicações (0-18); Eixo Y= total de citações recebidas por autor (0-762); Tamanho radio= GSRank (0-980); cor= autores).

No gráfico da Figura 6 a posição dos círculos que estão mais perto do Eixo Y são aqueles trabalhos que possuem menor tempo de publicação. Além disso, é preciso destacar que aquelas publicações que estão posicionadas mais perto do eixo Y, e que apresentam maior número de citações, coincidem com os trabalhos melhor posicionados no GSRank.

A Figura 6 mostra também o fato que das 7 pesquisas mais citadas e com mais de 500 citações recebidas (NK Hayles do 2011 e do ano 2007, L Manovich, R Kitchin, ML Jockers, P Juola e J Nerbonne) o 57% são das mais actuais, pois têm somente 6 anos de publicados, como é o caso das publicações de NK Hayles do 2011, R Kitchin, J Nerbonne e ML Jockers.

5 CONSIDERAÇÕES FINAIS

Foi possível concluir que o uso de softwares de visualização de dados nas análises métricas é comum na área da Ciência da Informação, embora, as configurações dos modelos complexos e altamente personalizados ainda sejam difíceis para os profissionais que não possuem as competências necessárias durante o desenvolvimento da pesquisa, o que demandaria uma quantidade significativa de tempo para correto domínio das ferramentas.

Também que, as principais publicações recuperadas sobre o tema estudado, não são oriundas de fontes ou tipologias documentais tradicionais de análise métrica da produção científica, e que os livros, são as principais fontes do foco de estudo.

Comportamentos interessantes foram identificados em relação às métricas de citações recebidas de artigos científicos de fontes alternativas às de sua publicação original, como é o caso das redes sociais acadêmicas (ex. academia.edu, ResearchGate). Este fato demonstra a tendência dos pesquisadores em compartilhar suas publicações e ganhar maior visibilidade para seu trabalho nestas plataformas, para além de sua publicação em periódicos científicos.

O ano de 2012 foi identificado como o ano mais produtivo em humanidades digitais e cujas publicações foram as mais citadas.

Identificou-se que a tendência geral é de que os documentos mais citados não apareçam nas 100 primeiras posições do ranking acadêmico do Google.

REFERÊNCIAS

ADAMS, J.L.; GUNN, K.B. Digital Humanities: Where to Start. **College & Research Libraries News**, v. 73, n. 9, p. 536-539, 569, 10 2012.

ATWOOD, T. P., e REZNIK-ZELLEN, R. Using the Visualization Software Evaluation Rubric to explore six freely available visualization applications. **Journal of eScience Librarianship**, v. 7, n. 1, e1122 p. 1-20, 2018. Recuperado de: <<https://escholarship.umassmed.edu/jeslib/vol7/iss1/3/>>. Acessado em: 15 novembro 2018

HARZING, Anne-Wil, et al. **The Publish or perish book**. Melbourne: Tarma software research, 2011.

MAURI, M., Elli, T., CAVIGLIA, G., UBOLDI, G., & AZZI, M. RAWGraphs: A Visualisation Platform to Create Open Outputs. In **Proceedings of the 12th Biannual Conference on Italian SIGCHI Chapter** (p. 28:1–28:5), 2017. New York, NY, USA: ACM. Recuperado de: <<https://doi.org/10.1145/3125571.3125585>>. Acessado em: 24 setembro 2018.

POOLE, A.H. The conceptual ecology of digital humanities. **Journal of Documentation**, Bradford, v. 73, n. 1, p. 91-122, 2017.

RAWGraphs Team. **Introduction to RAWGraphs**. Licensed under CC BY-NC-SA 4.0. Recuperado de: <<https://rawgraphs.io/learning/introduction-to-rawgraphs/>>. Acessado em: 23 setembro 2018.

VERBEKE, D. Digital Humanities and Libraries - DH 2014. **META: tijdschrift voor bibliotheek & archief**, v. 90, n. 7, p. 41-42, 2014.

WONG, S.H.R. Digital Humanities: What Can Libraries Offer? **Portal : Libraries and the Academy**, Baltimore, v. 16, n. 4, p. 669-690, 10 2016.

ORTEGA, N. R. Humanidades Digitales, Digital Art History y cultura artística: relaciones y desconexiones. **ARTNODES: Revista de Arte, Ciencia y Tecnología**, Catalunya, n. 13, p. 16-25, 2013.

RODRÍGUEZ-YUNTA, L. Humanidades digitales, ¿una mera etiqueta o un campo por el que deben apostar las ciencias de la documentación?. **Anuário ThinkEPI**, v. 7, pp. 37-43, 2013.