

ESTUDO MÉTRICO TEMÁTICO SOBRE BIBLIOTECA DIGITAL NO BRASIL: uma aplicação do software IRAMUTEQ

Márcio Henrique Wanderley Ferreira
Universidade Federal do Pernambuco, Recife, PE, Brasil
marcio.wferreira@gmail.com

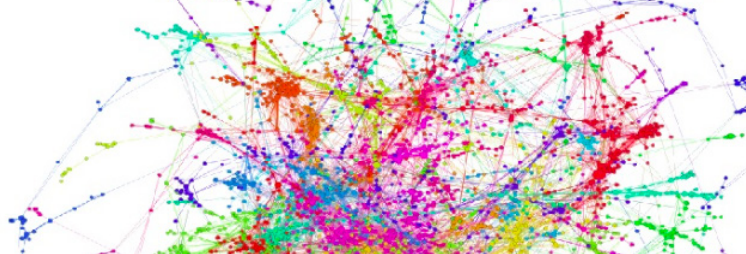
Renato Fernandes Corrêa
Universidade Federal do Pernambuco, Recife, PE, Brasil
fc_renato@yahoo.com.br

1 INTRODUÇÃO

De acordo com Harter (1996), a biblioteca digital contém representações digitais dos objetos nela encontrados e seu conceito teve origem nos anos de 1950 e vem evoluindo até hoje. Para a Digital Library Federation (1998 apud SAYÃO, 2009) as

Bibliotecas digitais são organizações que disponibilizam os recursos, incluindo pessoal especializado, para selecionar, estruturar, oferecer acesso intelectual, interpretar, distribuir, preservar a integridade e assegurar a persistência ao longo do tempo de coleções de trabalhos digitais, de forma que eles estejam pronta e economicamente disponíveis para uso de uma comunidade definida ou um conjunto de comunidades.

Nesse contexto, o objetivo deste trabalho é verificar o nível de proximidade e a frequência dos assuntos desenvolvidos conjuntamente com o descritor “Biblioteca Digital”. Para isso, utilizou-se o método estatístico de análise de frequência de termos buscando identificar e refletir sobre os principais conceitos que possuam relação com o tema, ao longo de 16 anos (2001 até 2017) de estudos desenvolvidos. O método escolhido levanta os temas mais relevantes que estão sendo desenvolvidos na Ciência da Informação (CI) sobre esta temática.



2 METODOLOGIA

O estudo adotou o método estatístico de análise textual. Inicialmente, buscou-se na BRAPCI o termo composto “biblioteca digital” no campo das palavras-chave. Ressalta-se que, para efeitos desta pesquisa, os resultados da busca pelo termo composto no singular englobam aqueles que se apresentam no plural. Ao realizar a busca que compreende o período de 1972 a 2017, em novembro de 2017, foram recuperados 100 trabalhos com a palavra-chave “Biblioteca digital”. Contudo, apenas 82 foram utilizados na pesquisa, por possuírem “Biblioteca digital” no campo das palavras-chave. Assim, foram realizadas as seguintes etapas: **1)** Coleta manual das palavras-chave, dos títulos e dos resumos dos campos dos 82 artigos para uma base de dados. Foram coletadas 430 palavras-chave, 82 resumos e 82 títulos; **2)** Formatação do arquivo de extensão (.txt) correspondente ao *corpus* da pesquisa, no formato padrão exigido pelo software IRAMUTEQ; **3)** Realização do pré-processamento textual, introduzindo a padronização dos termos e retirada (limpeza) daqueles que não tivessem relevância nos resultados.; **4)** Compilação dos dados e obtenção dos índices de frequência, ocorrência e de análise de similitude de palavras utilizando as seguintes ferramentas do software: aplicação da Lei de Zipf; cálculo de frequência de palavras; análise de similitude.

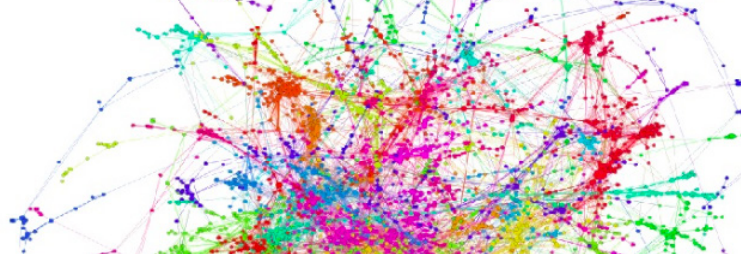
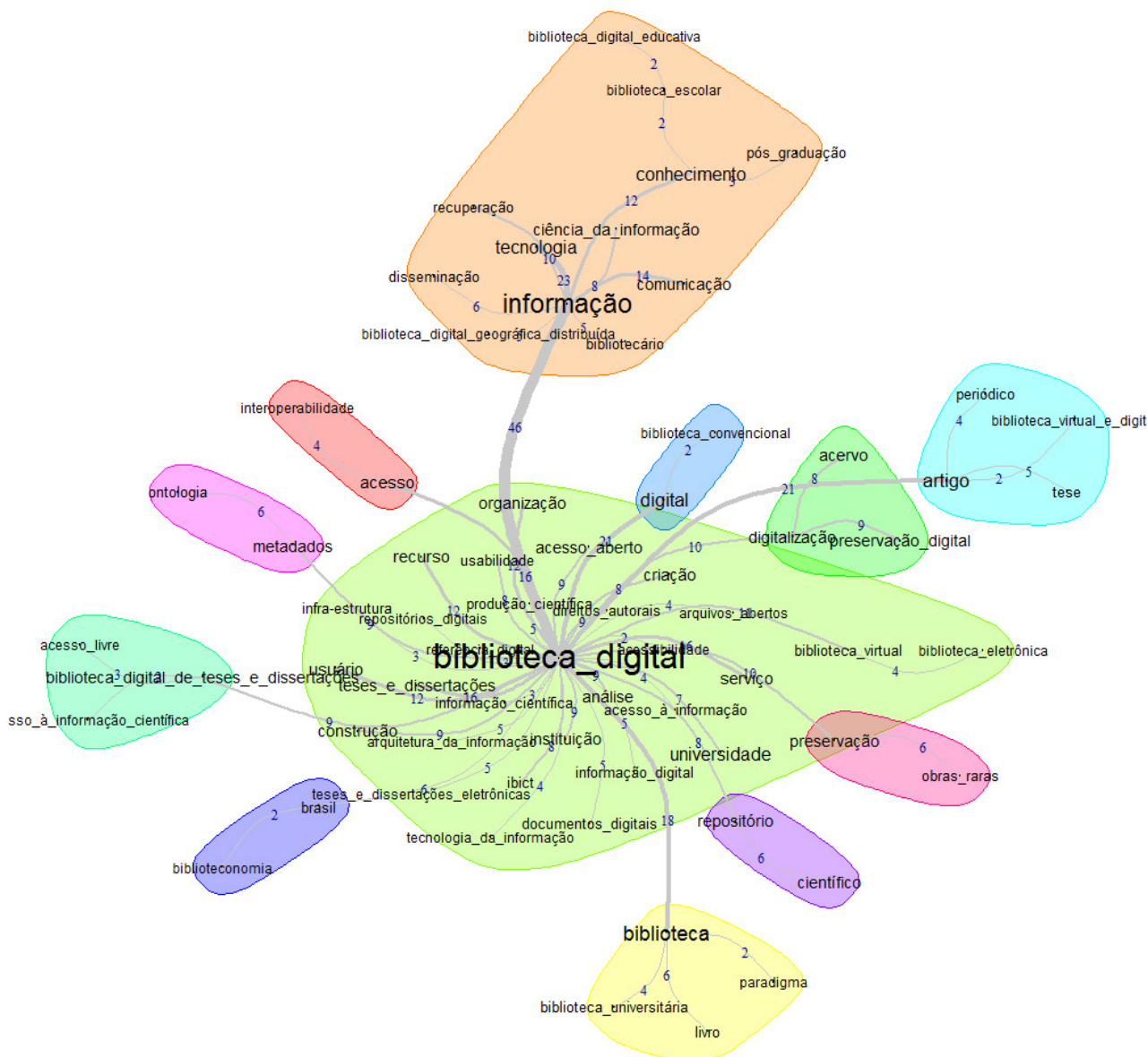
3 RESULTADOS

Primeiramente, optou-se por realizar uma análise que demonstrasse os índices de dispersão dos vocábulos analisados na pesquisa, dada pela Lei de Zipf. Nessa análise foi possível identificar que nos 82 artigos analisados existiam 12.440 ocorrências de palavras e 2.087 tipos de formas diferentes. O número do hápax demonstrou que 8,4% das ocorrências correspondem a 50,07% das formas, ou seja, cerca de metade dos termos correspondem a 91,6% das palavras em nível de ocorrência. Em seguida, foi feita uma análise de frequência, com o objetivo de identificar as palavras mais frequentes dentro do *corpus* observado. Pôde-se observar um conjunto de termos mais repetidos e também relevantes em termos semânticos. Assim, identificaram-se as seguintes palavras segui-



das do índice de repetições: biblioteca digital (203); informação (129); biblioteca (57); bdtd (26); teses e dissertações (25); preservação digital (24); ciência da informação (23); biblioteca virtual (16); documentos digitais (13) e informação digital (10).

Outro importante resultado é a análise de similitude com as principais temáticas encontradas nos artigos. Para realizar essa análise foram excluídas as seguintes classes gramaticais: artigos, advérbios, conjunções, preposições e verbos. Outro ponto a ser ressaltado, é que foram utilizados para essa amostra, os 70 vocábulos com maior índice de frequência e relevância temática, para os quais a frequência variou de 03 até 203. Esse método foi adotado porque permitiu uma melhor análise visual, vide Gráfico 1.

**GRÁFICO 1 - AGRUPAMENTOS DE TERMOS EM NÍVEL DE PROXIMIDADE E SIMILITUDE**

Fonte: Elaborado pelo autores, 2018.

No Gráfico 1, observa-se que os vocábulos (aqui chamamos de vértices) se ligam a outros por meio de arestas de cor cinza-claro. A espessura dessas “arestas” representa o grau de conexão entre os termos observados, assim também como os agrupamentos coloridos demonstram um grau de similitude e os valores numéricos em azul a quantidade de vezes em que os termos aparecem juntos no texto. Dessa maneira, podemos visualizar



que “biblioteca digital” se conecta com vértices como “biblioteca_virtual”, “informação_digital”, “usabilidade”, “arquitetura_da_informação”, “biblioteca_virtual” e “repositórios digitais”, todos no agrupamento em verde-claro (conforme o Gráfico 1). Enquanto que se conecta com outros agrupamentos de vértices da periferia (aglomerações de outras cores), e que possuem arestas de dimensão menores, os termos “biblioteca”, “repositório”, “preservação”, “artigo”, “digitalização”, dentre outros. Um ponto curioso nesse grafo, é que “informação” se conecta 46 vezes com “biblioteca_digital”, sendo o termo que mais se conecta com ele, porém não faz parte do agrupamento próximo à BD. Supõe-se que esses vocábulos em grupos mais distantes possuem algum grau de similitude com a temática central, porém entre si existem correlações mais fortes de adjacência textual que com as temáticas no grupo central. Assim, no instante em que “biblioteca_digital” está conectado com “informação”, percebe-se que os conceitos que envolvem os dois vocábulos possuem afinidade, mas que “informação” é tratada como um aspecto do tema principal independente de vocábulos em outros agrupamentos (contextos específicos).

4 CONSIDERAÇÕES

As análises realizadas demonstraram como o tema da “biblioteca digital” vem se conectando com outras temáticas importantes na CI, como biblioteca, preservação digital, acesso livre e os repositórios científicos, por exemplo. A metodologia deste trabalho envolveu a proposição e aplicação de um método e uma ferramenta para a realização de um estudo métrico temático, que consistiu na análise de temas ou assuntos presentes em registros bibliográficos. O IRAMUTEQ se apresentou como uma ferramenta útil para análise e visualização de informações textuais no desenvolvimento de estudos métricos temáticos. Mesmo com o uso do IRAMUTEQ, o método para realização do presente estudo métrico ainda requer grande esforço humano, principalmente para tratamento adequado dos termos compostos. Como sugestão de estudos futuros, pretende-se aprofundar nas análises estatísticas do método utilizado neste trabalho e aplicar outras



ferramentas, com fins de automação e/ou comparação na realização das etapas necessárias para alcance dos resultados.

AGRADECIMENTOS

Agradeço, primeiramente, à CAPES como promotora da minha bolsa de doutorado. Também externo agradecimentos ao PPGCI/UFPE e ao meu orientador Prof. Dr. Renato Fernandes Corrêa, que tem me apoiado em todos os sentidos.

REFERÊNCIAS

DIGITAL LIBRARY FEDERATION (Washington). **A working definition of digital library**. 2008. Disponível em: <<https://old.diglib.org/about/dldefinition.htm>>. Acesso em: 15 mar. 2018.

HARTER, S. P. What is a digital library? Definitions, content, and issues. In: INTERNATIONAL CONFERENCE ON DIGITAL LIBRARIES AND INFORMATION SERVICES FOR THE 21ST CENTURY, 1996. p. 8-17.

RATINAUD, P. IRaMuTeQ: Interface de R pour les Analyses Multidimensionnelles de Textes et de Questionnaires. **Téléchargeable à l'adresse: <http://www.iramuteq.org>**, 2009.

SAYÃO, L. F. Afinal, o que é biblioteca digital? **Revista USP**, São Paulo, v. 80, n. 1, p. 6- 17, fev. 2009.