

Acesso aberto ao conhecimento científico e as métricas informacionais: metainscrições científicas para identificar a organização intelectual em periódicos científicos

Alexandre Ribas Semeler (UFRGS) - alexandre.semeler@ufrgs.br

Veleida Ana Blank (UFRGS) - veleida.blank@ufrgs.br

Helen Beatriz Frota Rozados (UFRGS) - hrozados@gmail.com

Resumo:

Aborda os conceitos fundamentais sobre o acesso aberto ao conhecimento científico e as métricas informacionais. Propõe a análise métrica e visual da informação. Enfatiza a utilização da co-word analysis, conjuntamente aos métodos de visualização da informação, para o trato de metadados de publicações científicas Open Access (OA). Sugere a hipótese teórica: metadados de publicações científicas OA, de maneira semelhante aos artigos científicos, contêm informações textuais sobre determinado conteúdo científico (autor, instituição, títulos de artigos, resumos, palavras-chave, datas etc.) e, por essa razão, são consideradas metainscrições científicas portadoras de dados sobre as relações simbólicas e representam a organização intelectual do conteúdo científico dessas publicações. Conclui pela necessidade de se desenvolver uma metodologia alternativa para a análise métrica de publicações, principalmente para aquelas que estejam fora do núcleo ISI Thomson Scientific e que adotam a iniciativa de acesso aberto.

Palavras-chave: *Acesso aberto. Periódicos científicos. Métricas de informação. Metadados. Co-word analysis.*

Área temática: *Temática I: Tecnologias de informação e comunicação - um passo a frente*

**Acesso aberto ao conhecimento científico e as métricas informacionais:
metainscrições científicas para identificar a organização intelectual em periódicos
científicos**

Resumo:

Aborda os conceitos fundamentais sobre o acesso aberto ao conhecimento científico e as métricas informacionais. Propõe a análise métrica e visual da informação. Enfatiza a utilização da *co-word analysis*, conjuntamente aos métodos de visualização da informação, para o trato de metadados de publicações científicas *Open Access (OA)*. Sugere a hipótese teórica: metadados de publicações científicas *OA*, de maneira semelhante aos artigos científicos, contêm informações textuais sobre determinado conteúdo científico (autor, instituição, títulos de artigos, resumos, palavras-chave, datas etc.) e, por essa razão, são consideradas metainscrições científicas portadoras de dados sobre as relações simbólicas e representam a organização intelectual do conteúdo científico dessas publicações. Conclui pela necessidade de se desenvolver uma metodologia alternativa para a análise métrica de publicações, principalmente para aquelas que estejam fora do núcleo *ISI Thomson Scientific* e que adotam a iniciativa de acesso aberto.

Palavras-chave: Acesso aberto. Periódicos científicos. Métricas de informação. Metadados. *Co-word analysis*.

Área Temática: Temática I: Tecnologias de informação e comunicação – um passo a frente

1 INTRODUÇÃO

Os periódicos de acesso aberto e os repositórios institucionais amplificam e oferecem maior visibilidade para a Comunicação Científica na *web*. Estes consolidam um importante sistema de publicações fundamentadas no acesso aberto ao conhecimento científico. Talvez seja possível assegurar que o movimento *Open Access (OA)* revolucionou os processos comunicativos que envolvem o trabalho com informação científica em meio digital.

Sistemas de gerenciamento de informação científica, como *Open Journal System (OJS)*¹, representam, de maneira ampla e significativa, a estrutura tecnológica que dá suporte à produção, à comunicação e ao uso de periódicos científicos *OA* na Internet. Uma série de portais de acesso aberto a revistas científicas já foi implantado no País como os do IBICIT, da UNB, da UFRGS, da UFSC, da UEL entre outros. Assim, percebe-se que no Brasil e no mundo o *OJS* é amplamente utilizado. No entanto, apesar de diversas universidades e instituições científicas adotarem a ferramenta para gerenciar seus periódicos científicos *OA*,

¹ *Software open source* de gerenciamento e publicação de revistas eletrônicas foi desenvolvido e é disponibilizado pelo *Public Knowledge Project (PKP)*, fundado por John Willinsky, da *Faculty of Education at the University of British Columbia* desde 1998. A tradução para o português do Brasil é responsabilidade do IBICT e recebe o nome de Sistema Eletrônico de Editoração de Revistas (SEER).

até o momento são poucos os trabalhos que abordam o uso desse sistema de publicações aliado às perspectivas dos estudos métricos.

Os estudos métricos e visuais da informação buscam compreender as atividades exercidas por pesquisadores de universidades, institutos, laboratórios e/ou outras instituições científicas. Sendo assim, no que tange à justificativa de trabalho operacional busca-se, com este estudo, a proposição de novos serviços bibliotecários por meio da prática de métricas informacionais a respeito da pesquisa científica.

Nesse contexto, ressalta-se a *co-word analysis*, uma técnica métrica utilizada para analisar a ocorrência de palavras em textos; contudo, no estudo esta técnica é proposta para analisar metadados de periódicos científicos. Acredita-se que os campos de metadados como títulos, resumos e palavras-chave são importantes para identificação temática de tendências de pesquisa em publicações científicas OA. Isso, porque eles são inseridos pelos próprios autores no momento da submissão através do OJS tornando-se operadores de ligação entre os autores. Assim, se esses metadados representarem uma coleção, como os textos completos dos artigos, podem ser analisados e medidos para detectar as associações constantes entre os conceitos que delineiam determinadas áreas temáticas, campos ou disciplinas científicas.

Por esse fato, sugere-se a hipótese de que a *co-word analysis* possa ser utilizada como uma estratégia de análise de conteúdo em metadados de publicações científicas OA e não só para os artigos dessas publicações. Seguindo esse rumo, o objetivo geral do estudo apresentado nesse artigo é levantar fundamentos teóricos para circunscrever a análise métrica e visual de metadados de informação científica, disponível em publicações de acesso aberto, com vistas a incentivar o desenvolvimento de metodologia alternativa própria para análise métrica de publicações científicas.

2 FUNDAMENTAÇÃO TEORICA

A fim de introduzir questões-base sobre o movimento de acesso aberto ao conhecimento científico e as métricas, este ensaio aborda conceitos a respeito de publicações científicas AO e metadados.

A seguir abordam-se, resumidamente, as métricas informacionais (Biblio/Ciento/Info/Webometria), associadas aos métodos de visualização da informação em rede, com atenção especial ao método métrico da *co-word analysis*, que busca compreender o processo de coleta, extração e representação visual da informação textual que compõe metadados de publicações OA.

2.1 ACESSO ABERTO À INFORMAÇÃO CIENTÍFICA E METADADOS

O acesso livre à informação científica revela-se como uma nova tendência de estudos para Comunicação Científica. Inúmeras iniciativas consolidam as preocupações da comunidade científica com a filosofia *open*, ou seja, em direção ao uso de ferramentas, estratégias e metodologias para o desenvolvimento de aplicações de computador (*software livre*), arquivos abertos (*open archive*) e acesso aberto (*open access*). Essas novas iniciativas surgiram com a Internet e atualmente são muito influentes no que se refere aos recentes processos da Comunicação Científica (MUELLER, 2006).

O movimento de acesso aberto (*open access*) representa uma tendência atual dos estudos sobre a Comunicação Científica e surge como uma alternativa para divulgação e amplificação da visibilidade das publicações na *web*, revelando-se com um novo e fértil campo de estudos para pesquisas métricas.

Uma publicação científica *OA* define-se pela capacidade de estar disponível *online*, de forma gratuita e livre de direitos autorais ou restrições de licenciamento. O conceito de acesso aberto foi desenvolvido ao longo de três conferências: Budapeste, em 2001, Bethesda, em 2003 e em Berlim, em 2003. Nesses eventos, definiu-se que publicações *OA* são toda literatura disponibilizada de forma livre pela Internet e que essas devem permitir a qualquer utilizador ler, fazer *download*, copiar, distribuir e imprimir completamente o seu conteúdo. O objetivo do movimento *OA* é remover as barreiras entre todos os usos legítimos da literatura acadêmica (MUELLER, 2006, SUBER, 2004, BRKOVI, 2011, CRAWFORD, 2011).

Os principais veículos de comunicação e distribuição de publicações *OA* são as revistas científicas e os repositórios de acesso livre. Esses veículos podem ser classificados segundo duas vias: *green route* e *gold route*. Esses caminhos ou vias representam os tipos possíveis para sistemas de publicações *OA* (JEFFERY, 2006, HARNAD, 2009).

As estratégias da *green route*, ou via verde, propõem que universidades, institutos de pesquisa e outras instituições de cunho científico ou tecnológico criem seus próprios repositórios com o propósito de armazenar e dar maior visibilidade para a produção científica de seus pesquisadores. Nas palavras de Leite (2009, p. 17), “[...] a via verde significa o sinal verde dos editores científicos para o arquivamento da produção científica pelos autores em repositórios digitais.” Já a *gold route* propõe aos editores científicos a criação e/ou conversão das revistas científicas comerciais para revistas científicas de acesso livre. Conforme Leite

(2009, p. 7), "[...] a via dourada significa o acesso aberto promovido nos próprios periódicos científicos, de modo que os artigos científicos possam ser disseminados sem restrições de acesso ou uso."

Ao complementar, Crawford (2011, p. 15, tradução nossa) expande a definição de publicações *OA* explicando que "[...] contribuições *open access* incluem também os resultados originais de pesquisas científicas, dados brutos e metadados [...]". Nessa afirmação, o autor indica o termo metadados como uma perspectiva de estudos, isso porque o intercâmbio da informação científica em ambas as vias definidas como veículos de comunicação para publicações *OA* se dá por meio de metadados.

O protocolo de metadados utilizado pelo *Open Journal System* é o *Open Archives Initiative Protocol for Metadata Harvesting (PMH-OAI)*, padrão de interoperabilidade e coleta de metadados que busca alcançar o princípio da interoperabilidade entre metadados de sistemas digitais de publicações científicas *OA*. No âmbito tecnológico, esse protocolo estabelece os princípios e as técnicas para integração de arquivos abertos. Ele é fruto de um encontro dos representantes de organizações que gerenciam serviços de arquivos abertos (*eprints*) realizado em Santa Fé, Novo México, em 1999 (CAPLAN, 2003).

Entende-se por *Harvesting* – coleta de metadados – a extração de descrições de documentos em rede. Geralmente, os documentos em revistas científicas ou em repositórios de acesso livre são acompanhados de informações auxiliares que vão além do pesquisador, do tempo e do lugar em que o experimento foi conduzido. Esta informação auxiliar constitui o que esse estudo entende por metadados, ou seja, um conjunto de dados sobre pesquisas científicas que podem estar acoplados, embutidos ou diretamente associados às publicações digitais em bases de dados (CAPLAN, 2003).

Metadados de publicações científicas *OA* são um dos focos temáticos deste ensaio. Neste sentido, surge a hipótese de que os metadados oriundos de publicações científicas *OA* podem gerar dados sobre as pesquisas realizadas por uma comunidade científica em particular. Dessa forma, o próximo tópico pretende clarificar o conceito de metadados, com base na hipótese de que metadados podem servir de *corpus* para a análise métrica e visual em publicações científicas *OA*.

Ao definir o termo metadados, Caplan (2003, p. 10, tradução nossa) argumenta que “[...] o termo refere-se a todos os dados utilizados para auxiliar na identificação, descrição e localização de recursos eletrônicos em rede [...]”. A autora (2003, p. 14, tradução nossa) explica que “[...] o prefixo meta é comumente usado para significar algo sobre algo, por exemplo, uma metalinguagem usada para descrever outra linguagem [...]”. São informações

que se encontram acopladas, embutidas ou diretamente associadas na maioria dos objetos digitais. Servem para descrição, administração e mapeamento da estrutura e do conteúdo desses objetos digitais. São similares, por exemplo, a um catálogo de fichas de uma biblioteca (CAPLAN, 2003, MOORE; MOREAU, 2001).

Metadados são, muitas vezes, definidos como dados sobre dados ou informação sobre informação, então, metadados científicos são dados sobre as ciências. Os tipos de metadados são três: metadados descritivos, aqueles que descrevem um recurso para fins como a descoberta e a identificação; metadados estruturais, que indicam como objetos são compostos e colocados juntos; e metadados administrativos, que provêm informações de auxílio para gerenciar recursos sobre a gestão técnica, de direitos e de preservação de metadados (CAPLAN, 2003, NISO, 2004).

Um grande número de propostas de padronização de metadados tem sido desenvolvido em âmbito internacional: *FGDC (Federal Geographic Data Committee)*, *IMS (Instructional Management Systems)*, *LOM – IEEE (Learning Object Meta data)*, *VRA (Visual Resources Association)*, entre outros. Diferentes comunidades e domínios do conhecimento propõem diferentes padrões de metadados com a intenção de fornecer orientações quanto à estrutura e ao conteúdo de grandes massas de dados em áreas específicas do conhecimento. O *Dublin Core* é um padrão de metadados utilizado para descrever publicações científicas *open access*, composto por 15 elementos, planejado para facilitar a descrição de recursos eletrônicos (CAPLAN, 2003, ZENG; QIN, 2008),

Segundo o protocolo *Harvesting*, os metadados de publicações científicas *OA* adotam as especificações *Dublin Core*, por isso são fonte essencial para a avaliação métrica. Teoriza-se, na sequência, sobre o conceito de análise métrica da informação fundamentado na Bibliometria, Cientometria, Infometria e Webometria. No entanto, ressalta-se que o objetivo aqui não é discutir tais áreas em si, mas introduzi-las como fundamentos teóricos indispensáveis para análise métrica de informação, contida em metadados científicos.

2.2 MÉTRICAS DA INFORMAÇÃO

Podemos compreender a análise métrica da informação como um princípio que tem como base a medição da informação segundo ferramentas, métodos e técnicas provenientes da Bibliometria, da Cientometria, da Infometria e da Webometria. Conforme explica Gonzáles-Sae (2005, p. 32), “[...] todas têm funções semelhantes, mas, ao mesmo tempo, cada uma

delas se propõe medir a difusão do conhecimento científico e o fluxo da informação sob enfoques diferentes”.

A Bibliometria preocupa-se em quantificar os processos de comunicação escrita, conforme a natureza e a evolução das disciplinas, no que essas se refletem na literatura científica. A Cientometria aplica as mesmas técnicas, mas se restringe a analisar a atividade de produtividade intelectual de pesquisadores, busca identificar as leis que regem as ciências com o uso de ferramentas específicas como, por exemplo, os mapas das ciências. Essa métrica procura compreender os limites, as variações e as relações de influência que existem em uma comunidade científica (GONZÁLES-SAE, 2005, FARIA, 2005, VANTI, 2007, ALVARADO, 2007).

A Infometria propõe-se a estudar a informação em qualquer formato, seja ela científica ou não. O escopo da investigação infométrica é mais abrangente do que aqueles feitos pela Cientometria e pela Bibliometria, pois pesquisa as modalidades de produção, comunicação e uso de informação em comunidades acadêmicas e não acadêmicas (GONZÁLES-SAE, 2005, FARIA, 2005, VANTI, 2007, ALVARADO, 2007).

Desse modo, a Infometria fundamenta-se na combinação de técnicas de recuperação da informação e na medição dos fluxos informacionais em bases de dados. A última, uma área em constante expansão, configura a Webometria a qual comporta os estudos sobre os aspectos quantitativos da construção e do uso dos recursos de informação, estruturas e tecnologias na *web*, utilizando enfoques bibliométricos e infométricos para análise de informação (GONZÁLES-SAE, 2005, FARIA, 2005, VANTI, 2007, ALVARADO, 2007).

As principais atividades de investigação sobre a análise métrica da informação em publicações científicas são originadas de métodos cientométricos fundados na Bibliometria. Por exemplo, duas medidas básicas da Bibliometria: frequência e coocorrência podem ser utilizadas para o desenvolvimento de indicadores quantitativos sobre ciência e tecnologia.

A definição de frequência pode ser compreendida pela constância na qual as palavras-chave de uma coleção de documentos aparecem juntas. Ela é usada como medida-base para representar a coocorrência. Os agrupamentos sucessivos das palavras por meio da quantificação da sua frequência permitem a visualização da coocorrência (NEFF; CORLEY, 2009).

Esses agrupamentos de frequência podem assumir diversos níveis de agregação. Explica Leydesdorff (2001, p. 14, tradução nossa) que "[...] as palavras organizam os textos, artigos científicos os periódicos, periódicos os diretórios de periódicos, cientistas compõem grupos de pesquisa, grupos de pesquisa às comunidades [...]" e assim, sucessivamente. Pode-

se dizer que essas agregações representam possíveis unidades de análise para um estudo sobre coocorrências cientométricas fundamentadas na frequência.

Geralmente, a frequência de coocorrência de uma palavra em um, dois ou mais textos indica as relações entre eles. A frequência de uma única palavra em vários textos pode resultar na construção de redes de palavras que relacionam os seus autores em níveis sociocognitivos (KOSTOFF, 2003).

Já a *co-word analysis* é uma das técnicas de análise cientométrica, fundamentada na frequência da coocorrência de palavras em textos científicos, conforme será explanado.

2.3 CO-WORD ANALYSIS

A *co-word analysis* possui suas origens na linguística, na lexicografia e na linguística computacional. Os primeiros estudos foram sobre a coocorrência de palavras em textos não científicos, bem como seus significados e as razões para duas palavras coocorrem no mesmo contexto. Ressalta-se que o interesse da linguística computacional centrou-se na recuperação de informação, na lexicografia mediada por computador e em modelos de armazenamento da linguagem (KOSTOFF, 2003).

Os primeiros estudos sobre a *co-word analysis* foram úteis para mostrar as relações entre as palavras e os termos que podem ser extraídos e quantificados em um texto. No entanto, não foram aplicados para avaliar as tendências de pesquisa ou para gerar subsídios de apoio a políticas de investigação científica. Assim, deixam de lado o potencial da *co-word analysis* em identificar as tendências de pesquisa em grandes massas de informação textual científica, pois nada impede que se considere a produção científica como um gênero literário específico e se aplique a *co-word analysis* a ele (KOSTOFF, 2003).

A primeira vez em que a *co-word analysis* foi utilizada para avaliação da pesquisa científica foi na proposta feita por Michael Collon junto ao *Centre de Sociologie de L'Innovation* na *École Nationale Supérieure des Mines* de Paris, em 1983. O autor acreditava que, por meio do método de *co-word analysis*, seria possível identificar redes e estudar sua evolução com base na análise de documentos. Seu objeto de estudo foram as palavras-chave de artigos científicos. O autor não limita o uso da técnica apenas às publicações científicas. Nesse trabalho, Collon não analisou os textos científicos, mas as palavras-chave desses, realizando, assim, uma das primeiras experiências com metadados de publicações científicas (CALLON; COURTIAL; WILLIAM, 1983).

No entanto, os especialistas advertem que a expressão mais codificada da pesquisa científica é o texto em artigos científicos, pois eles contêm informações relevantes sobre diversos aspectos da organização da pesquisa, fornecendo aos estudos da ciência uma ferramenta promissora de investigação no cruzamento de diferentes dimensões analíticas (DE BELLIS, 2009).

A *co-word analysis* fundamenta-se na coocorrência e envolve uma série de etapas no que tange a sua operacionalização. Para De Bellis (2009, p. 143, tradução nossa), são três: “[...] extração, classificação e agrupamento e visualização da *co-word* [...]”. Já para Yang; Wu; Cui (2011, p. 660, tradução nossa), são cinco as etapas necessárias para realização da *co-word analysis* quando se pretende a visualização da informação “[...] coleta de literaturas, [...] extração de dados, [...] processamento de dados, [...] análise de dados, [...] visualização dos resultados [...]”. Mesmo não concordando com o número de etapas para realização da *co-word analysis*, os autores consolidam os procedimentos necessários para utilização da técnica.

Assim, as etapas de operacionalização da *co-word analysis* podem ser descritas, inicialmente, como: **extração**, proposta por De Bellis, que circunscreve as duas primeiras etapas propostas por Yang, Wu e Cui (2011), **coleta de literaturas** e **extração de dados**. Segundo De Bellis (2009, p. 143, tradução nossa), “[...] dado um corpus de texto, normalmente um conjunto de documentos científicos, a frequência de coocorrência de todos os pares de palavras significativas deve ser extraída e calculada.” Ou seja, nesse momento da *co-word analysis*, são realizadas pesquisas e coletas de dados sobre literatura científica. A coleta de dados exige que se extraia um grupo de palavras representativas como objeto de estudo. A extração gera informação quantitativa sobre os padrões de frequência dos pares de palavras.

Durante esse processo, surge a necessidade de normalizar os dados coletados. Chamada por De Bellis (2009) de **classificação**, esta fase compreende o que Yang, Wu e Cui (2011) denominam de **processamento e análise dos dados**. A classificação, nesse contexto, consolida-se por meio do processamento e da análise de dados. Conforme explica De Bellis (2009, p. 144, tradução nossa), “[...] uma vez que todos os pares de palavras forem contados, os valores dispostos em uma matriz simétrica de coocorrência, uma variedade de manipulações pode ser feita [...]”. A multiplicidade de análises que podem ser aplicadas exige que se definam critérios de classificação para o uso das palavras, por exemplo: classificação, inclusão, proximidade e equivalência. Esses critérios definem a força relativa entre as ligações de coocorrência das palavras. No caso mais geral, determinam a hierarquia

dos problemas de pesquisa e as áreas que potencialmente crescem em um domínio científico representado na coleção analisada (DE BELLIS, 2009, YANG; WU ; CUI, 2011).

Existem muitos **métodos de visualização da informação** que podem ser aplicados na *co-word analysis*. Os três métodos de visualização mais utilizados são os seguintes: *cluster tree*, *strategy diagram*, *social network maps*. Esses três métodos possuem características próprias. O primeiro, *cluster tree*, mostra as estruturas de um assunto. A *strategy diagram* revela a importância de temas tópicos na estrutura. E a *social network maps* é utilizada para interpretar o relacionamento interno entre os temas. A integração dos resultados de visualização de diferentes métodos pode auxiliar na obtenção dos resultados de pesquisas cientométricas (YANG; WU; CUI, 2011).

A combinação entre as métricas e os métodos de visualização da informação consolidam uma boa estratégia para elaboração dos resultados de uma pesquisa cientométrica a respeito de metadados em publicações científicas OA.

2.4 ANÁLISE VISUAL DA INFORMAÇÃO

A visualização da informação é a forma pela qual se pode representar a informação de maneira gráfica ou visual. Seu propósito é facilitar a assimilação e o entendimento da informação. Surge da necessidade de se assimilarem grandes quantidades de informação. É tarefa do comunicador que deve transformar dados abstratos em mensagens visíveis. As ferramentas mais comuns para aplicação e o uso prático de tal técnica de representação são os mapas conceituais e os diagramas de rede (PINTO *et al.*, 2009).

A visualização da informação é de natureza interdisciplinar e envolve a Computação Gráfica, a Geografia e a Ciência da Informação, possui um grande potencial para melhorar a forma de acesso, processamento e gerenciamento de grandes quantidades de informação. Sendo assim, o papel dessa técnica é o de reduzir os dados a um único ambiente, simplificando a análise destes. Dentre as ferramentas utilizadas por esse tipo de representação estão os mapas das ciências (BÖRNER; CHEN, 2002).

Nas palavras de Börner (2010, p. 8, tradução nossa), “[...] os mapas das ciências têm guiado nossa busca por conhecimento, permitindo-nos visualizar os resultados científicos, auxiliando na navegação, na compreensão e na comunicação das mudanças ocorridas nas ciências e na tecnologia.” As representações visuais dos resultados de análises cientométricas constituem a *cienciografia*.

Com o objetivo de fornecer subsídios para elaboração e interpretação de mapas das ciências, a cienciografia funciona como ferramenta para visualização das relações e múltiplas perspectivas de uma comunidade científica. Por meio de seis etapas é possível elaborar um cienciograma: obtenção de dados, definição da unidade de estudo (autores, instituições ou palavras-chave), seleção das medidas de análise, cálculo de similaridade entre as unidades (correlação de dados), reordenação das unidades de análise e visualização gráfica para análise e interpretação (PINTO *et al.*, 2009).

O mapeamento das relações existentes em um campo de pesquisa científica pode ser realizado por meio de análise quantitativa sobre a ocorrência de dados textuais ou de registros bibliográficos sobre determinada produção científica. A base para o mapeamento semântico expressa-se com a elaboração de matrizes de coocorrências de palavras.

Assim, conclui-se que um mapa da ciência é uma representação espacial das relações entre disciplinas, campos, especialidades, artigos e/ou autores, fornecendo a descrição da estrutura intelectual de uma área de pesquisa, pois facilita a busca e a recuperação de informações em grandes conjuntos de dados (DE BELLIS, 2008).

Nesse sentido, surge a necessidade do desenvolvimento de um aparato metodológico para tratar do processo de elaboração e interpretação de dados visuais gerados por meio do mapeamento cientométrico.

3 CONSIDERAÇÕES FINAIS

Nesse texto foram elencados alguns dos conceitos fundamentais para o acesso aberto ao conhecimento científico e as métricas informacionais: publicações *OA*, metadados, análise métrica e visual da informação, enfatizando-se a *co-word analysis* em conjunto com os métodos de visualização da informação em metadados de publicações científicas *OA*.

Pelo estudo de tais conceitos e técnicas chega-se a seguinte hipótese teórica: metadados de publicações científicas *OA*, de maneira semelhante aos artigos científicos, contêm informações textuais sobre determinado conteúdo científico (autor, instituição, títulos de artigos, resumos, palavras-chave, datas etc.) e, por essa razão, são metainscrições científicas. Essas informações, como as palavras inscritas nos artigos científicos, são portadoras de dados sobre as relações simbólicas e representam a organização intelectual do conteúdo científico dessas publicações.

Diferentemente do texto completo de um artigo científico, um arquivo de metadados em uma base de dados representa um conjunto de fragmentos de vários textos científicos,

como os títulos, os resumos e as palavras-chave. Essas informações são dados fragmentados sobre o conhecimento registrado por uma comunidade de autores e instituições de pesquisa em uma base de dados. Além de trazerem consigo uma coletividade de dados sobre o saber científico, os metadados tratam de informações a respeito do plano da recuperação, do processamento, da preservação e da gestão de informação científica em meio digital. Nas palavras de Caplan (2003, p. 76) “[...] são um conjunto de informação estruturada sobre recursos de informação em qualquer mídia ou formato”. Eles são o objeto de estudo ideal para a coleta e a análise cientométrica.

Atualmente, os metadados de referência e os serviços de citação utilizados para análise métrica de publicações científicas são extraídos de grandes bases de dados como a do *Institute for Scientific Information (ISI)*². A maioria dos indicadores cientométricos (fator de impacto de periódicos, citações, cocitações etc.) é oriunda de metodologias métricas que fazem uso dessa base de dados ou de outras de origem proprietária – como a *SCOPUS*. Mueller postula que (2006, p. 30) o *ISI* praticamente é “[...] a única fonte universalmente ‘legitimada’ [...] para estabelecer não apenas os dados de citação e indicadores, mas também as fórmulas utilizadas para calcular tais indicadores.”.

Essa prática torna necessária a proposição de soluções alternativas para a análise cientométrica de publicações, principalmente para aquelas que estejam fora do núcleo *ISI Thomson Scientific* e adotem a iniciativa de acesso aberto. Tal constatação revela a da proposição de um método para a avaliação métrica e visual de metadados de publicações científicas *OA* que utilizem o *Open Journal System (OJS)*. Acredita-se que a investigação métrica aplicada ao estudo de publicações *OA* poderá fornecer novos meios para avaliar, prever e analisar o conhecimento científico armazenado em periódicos científicos e repositórios institucionais de acesso livre.

Assim, cabe buscar aplicações práticas para tais conceitos. Em um entendimento amplo, sugere-se o desenvolvimento de uma metodologia para análise métrica e visual de informação científica em publicações de acesso aberto. A investida metodológica deverá partir em direção à identificação de tendências temáticas contidas em metadados de publicações científicas *OA*. Por meio da exploração de *softwares* específicos, deverá buscar ferramentas e recursos tecnológicos que propiciem a pesquisa quantitativa e a representação visual de informações contidas em metadados.

Conclui-se, portanto, que é preciso desenvolver um método que busque a instrumentação das etapas da extração e coleta de metadados; normalização dos dados do

OJS; elaboração de lista de palavras; análise da frequência de palavras; análise das coocorrências de palavras; criação de matrizes de coocorrência; elaboração de representações visuais – mapas de redes; e, por fim, interpretação dos resultados.

No entanto, não cabe aqui desenvolver tal método, mas sim alertar para sua importância e lançar a ideia para seu desenvolvimento em estudos na área.

OPEN ACCESS TO SCIENTIFIC KNOWLEDGE AND INFORMATIONAL METRICS: meta-scientific inscriptions to identify the intellectual organization in scientific journals

Abstract: It addresses the basic concepts of open access to scientific knowledge and informational metrics. It proposes a metric and visual analysis of the information. Emphasizes the use of co-word analysis, together with information visualization methods for the treatment of metadata of Open Access (OA) scientific publications. Suggests a theoretical hypothesis: metadata of OA scientific publications, similar to scientific articles, contain textual information about a particular scientific content (author, institution, article titles, abstracts, keywords, dates etc..) and, therefore, are considered meta-inscriptions carrying scientific data about the symbolic relationships and represent the intellectual organization of the scientific content of these publications. Concludes with the need to develop an alternative methodology for metric analysis of publications, particularly those who are outside the core ISI Thomson Scientific and adopt open access initiative.

Keywords: Open access. Scientific journals. Metrics information. Metadata. Co-word analysis.

REFERÊNCIAS

ALVARADO, R. A. Bibliometria: história, legitimação e estrutura. In: TOUTAIN, L. M. **Para entender a Ciência da Informação**. Salvador: EDUFBA, 2007. p. 185-217.

BÖRNER, K. **Atlas of Science: visualizing what we know**. London: MIT Press, 2010.

BÖRNER, K.; CHEN, C. Visual Interfaces to Digital Libraries: motivation, utilization, and socio-technical challenges. **Lecture Notes in Computer Science**, v. 2539, n. 2539, p. 1-9, 2002.

BRKOVI, M. Academic Authors, Scientific Information and Open Access Publishing. In: KATSIRIKOU, A. **Open Access to STM Information: Trends, Models and Strategies for Libraries**. Boston: International Federation of Library Associations, 2011.

CALLON, M.; COURTIAL, J.; WILLIAM, A. A from translations to problematic networks: an introduction to co-word analysis. **Social Science Information**, v. 22, n. 191, 1983.

CAPLAN, P. **Metadata fundamentals for all librarians**. La Vergne: Lightning Source, 2003

CRAWFORD, W. **Open access: what you need to know now**. Chicago: ALA Editions, 2011.

² http://thomsonreuters.com/products_services/science/science_products/a-z/science_citation_index/

DE BELLIS, N. **Bibliometrics and citation analysis**: from the Science Citation Index to Cybermetrics. Lanham, Md.: Scarecrow Press, 2009.

FARIA, L. **Prospecção tecnológica em materiais**: aumento da eficiência do tratamento bibliométrico. Aplicação na análise de tratamentos de superfície resistentes ao desgaste. São Carlos, 2005. 187f. Tese (Doutorado em Ciência e Engenharia de Materiais) – Universidade Federal de São Carlos, 2005.

GONZÁLES-SAE, S. **Aplicação de métodos bibliométricos e da co-words analysis na avaliação da literatura científica brasileira em ciências da saúde de 1992 a 2002**. São Paulo, 2005, 215f. Tese (Doutorado em Ciência e Engenharia de Materiais) – Universidade Federal de São Paulo, 2005.

HARNAD, S. Open access scientometrics and the UK research assessment exercise. **Scientometrics**, v. 79, n.1, 2009.

JEFFERY, T. Open access: an introduction. **ERCIM News**, n. 63, 2006. Disponível em: <http://www.ercim.eu/publication/Ercim_News/enw64/jeffery.html>. Acesso em: 15 maio 2012.

KOSTOFF, R. **Science and technology text mining**: origins of database tomography and multi-word phrase clustering. 2003. Disponível em: <<http://handle.dtic.mil/100.2/ADA416268>>. Acesso em: 15 maio 2012.

LEITE, F. C. L. **Como gerenciar e ampliar a visibilidade da informação científica brasileira**: repositórios institucionais de acesso aberto. Brasília, DF: IBICT, 2009.

LEYDESDORFF, L. **The challenge of scientometrics**: the development, measurement, and self-organization of scientific. Amsterdam: DSWO Press, 2001.

MUELLER, Suzana Pinheiro Machado. A comunicação científica e o movimento de acesso livre ao conhecimento. **Ciência da Informação**, v.35, n.2, p. 27-38, 2006. Disponível em: <<http://www.ibict.br/cienciadainformacao/viewarticle.php?id=925&layout=abstract>>. Acesso em: 3 jul. 2012 .

MOORE, G.; MOREAU, L. **From metadata to links**. The Sixth Workshop on Open Hypermedia Systems (OHS 6). Berlin: Springer-Verlag, 2001. Disponível em: <<http://eprints.soton.ac.uk/id/eprint/254475>>. Acesso em: 15 maio 2012.

NISO - NATIONAL INFORMATION STANDARDS ORGANIZATION. **Understanding metadata**. Bethesda: NISO, 2004.

NEFF, M.W.; CORLEY, E. A. 35 Years and 160,000 Articles: A Bibliometric exploration of the evolution of ecology. **Scientometrics**, v.80, n.3, p.657-682, 2009.

PINTO, A. L.; BARQUÍM, B. A. R.; EFRAIN-GARCIA, P.; GONZÁLES, J. A.M. Visualização da informação das redes sociais através de programas de cienciografia. In: POBLACION, D. A.; MUGNAINI, R. **Redes sociais e colaborativas em informação científica**. São Paulo: Angellara, 2009.

SUBER, P. **A very brief introduction to Open Access**. 2004. Disponível em: <<http://www.earlham.edu/~peters/fos/brief.htm>>. Acesso em: 15 maio 2012.

VANTI, N. **Links hipertextuais na comunicação científica**: análise webométrica dos sítios acadêmicos latino-americanos em Ciências Sociais. Porto Alegre, 2007. 292 f. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul. Porto Alegre, 2007.

XXV Congresso Brasileiro de Biblioteconomia, Documento e Ciência da Informação –
Florianópolis, SC, Brasil, 07 a 10 de julho de 2013

YANG, Y.; WU, M.; CUI, L. Integration of three visualization methods based on co-word analysis.
Scientometrics, v.90, n.2, p. 659-673, 2011.