

XIX encontro nacional
de pesquisa em
ENANCIB ciência da informação

// SUJEITO INFORMACIONAL E AS
PERSPECTIVAS ATUAIS EM CIÊNCIA
DA INFORMAÇÃO. //

22-26
OUTUBRO
2018
LONDRINA/PR



XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018

GT-7 – Produção e Comunicação da Informação em Ciência, Tecnologia & Inovação

UMA ANÁLISE TEMPORAL DOS PRINCIPAIS TÓPICOS DE PESQUISAS EM CIÊNCIAS SOCIAIS APLICADAS NO BRASIL

Jether Gomes (CEFET-MG)

Thiago M. R. Dias (CEFET-MG)

Gray F. Moita (CEFET-MG)

A TEMPORAL ANALYSIS OF THE MAIN TOPICS OF RESEARCH IN SOCIAL SCIENCES APPLIED IN BRAZIL

Modalidade da Apresentação: Comunicação Oral

Resumo: A ciência apresentou um crescimento impressionante nas últimas décadas quanto ao número de publicações de artigos científicos, sendo que, os fatores essenciais para esse desenvolvimento são o grande número de informações disponibilizadas na internet e a ampliação da sociabilização entre os pesquisadores. Consequentemente, nota-se um esforço global de todas as áreas do conhecimento, negócios e fundações interessados em entender o patamar científico das nações ou grupos individualizados para, por exemplo, servir de base na construção de políticas científicas visando novos avanços na ciência ou impulsionar grupos de pesquisas a se tornarem mais produtivos. Neste contexto, o objetivo deste trabalho é identificar e analisar temporalmente os principais tópicos de pesquisa publicados durante a trajetória da ciência brasileira por pesquisadores doutores que possuem currículos cadastrados na Plataforma Lattes e, que se dedicam a grande área do conhecimento Ciências Sociais aplicadas. Para isso, após a aquisição dos currículos, foi desenvolvido um arcabouço de componentes responsável por filtrar, tratar e consolidar separadamente os dados das publicações científicas a serem analisados por métodos bibliométricos. Dentre estes dados, estão todas as palavras-chave de artigos publicados em periódicos entre 1997 e 2016, que são ranqueadas através da aplicação de uma medida de popularidade de tópico (TF.FI) proposta, que considera tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados. Os resultados possibilitaram a identificação e uma análise temporal dos principais tópicos investigados pelos doutores a cada biênio. Como considerações, observa-se que os doutores brasileiros que atuam na grande área de Ciências Sociais Aplicadas dedicaram mais esforços a um conjunto reduzido de tópicos de pesquisas, e, que a troca dos principais interesses foi

evoluindo gradualmente ao longo do tempo, possivelmente devido uma demanda da sociedade ou por maturação de um determinado tema.

Palavras-Chave: Palavras-chave; Tópicos de Pesquisas; Análises Bibliométricas; Plataforma Lattes.

Abstract: Science has shown an impressive growth in the last decades in the number of publications of scientific articles, and the essential factors for this development are the large number of information available on the Internet and the expansion of socialization among researchers. As a result, there is a global effort of all areas of knowledge, business and foundations interested in understanding the scientific level of nations or individualized groups, for example, to serve as the basis for the construction of scientific policies aiming at new advances in science or to promote groups to become more productive. In this context, the objective of this work is to identify and analyze the main topics of research published during the course of Brazilian science by PhD researchers who have curricula registered in the Lattes Platform and who are dedicated to a large field of applied social sciences. For that, after the curricula was acquired, a framework of components was developed responsible for filtering, treating and consolidating separately the data of the scientific publications to be analyzed by bibliometric methods. Among these data are all the keywords of articles published in journals between 1997 and 2016, which are ranked through the application of a proposed topic popularity measure (TF.FI), which considers both the quantitative characteristics of the keywords of the articles and qualitative of the journals in which such articles were published. The results allowed the identification and a temporal analysis of the main topics investigated by the PhDs every two years. As final considerations, it is observed that Brazilian PhDs working in the large area of Applied Social Sciences have devoted more effort to a reduced set of research topics, and that the exchange of major interests has gradually evolved over time, possibly due to a demand of the society's demand or maturity of a particular topic.

Keywords: Keywords; Research Topics; Bibliometric Analysis; Lattes Platform.

1 INTRODUÇÃO

A ciência apresentou um crescimento impressionante nas últimas décadas quanto ao número de registros de publicações de artigos científicos (RIDKER e RIFAI, 2013). Tal crescimento é verificado através resultados de esforços coletivos de pesquisadores de todos os domínios ao redor do mundo. Embora essa organização social tenha existido em épocas anteriores, a internet e seus serviços, é quem fornece a base principal para sua ampliação (CASTELLS, 2010).

Para Ferreira (2010), a comunidade científica tem a necessidade e o compromisso de tornar público todo o conhecimento produzido durante o desenvolvimento de suas pesquisas. A divulgação deste conhecimento tem ajudado a compreender melhor o mundo, e, com isso, permite o desenvolvimento da sociedade e de padrões para a vida humana.

Num âmbito mais específico, entender o avanço da produção científica é de suma importância para aumentar o poder da competitividade em negócios (SAES, 2005; CAVACINI, 2016; CHEN et. al. 2015). Em Brito et. al. (2016), os autores destacam que trabalhos desta natureza são considerados urgentes no Brasil e podem retratar o que é desenvolvido e

publicado em ciência, possibilitando gerar parâmetros para nortear esforços e investimentos com intuito de impulsionar resultados de pesquisa.

Diante disso, pesquisadores de todos os domínios têm dedicado esforços para analisar a produção científica sobre diferentes perspectivas, como, por exemplo, através de análise de tópicos de pesquisas. Nesse caso, um tópico pode ser entendido como termo que representa um dos assuntos associados a um determinado documento (BORGES et. al., 2015).

Assim, esforços para analisar tópicos constituem formas de aprimorar a compreensão do que tem sido produzido acerca da ciência. Adicionalmente, outras informações podem ser descobertas, a saber: (1) principais tópicos em discussão, (2) co-ocorrência dos tópicos de pesquisas, (3) tópicos de pesquisas abordados por áreas diferentes, (4) revisão da literatura e, (5) indicação de tendências de pesquisas. A partir destas descobertas, ainda é possível realizar análises baseadas em agrupamento, ranqueamento e recomendação.

Estes trabalhos geralmente exploram repositórios de artigos científicos, analisando seus títulos e resumos para extrair tópicos de pesquisas e analisá-los através de métodos bibliométricos ou técnicas de análises de redes sociais ou análises de tendências. Entretanto, títulos e resumos de trabalhos podem não representar em completude os assuntos abordados, devido a necessidade dos autores destes trabalhos se preocuparem com a semântica e estrutura dos termos.

Logo, uma abordagem alternativa é a análise das palavras-chave de artigos científicos, pois são inseridas por seus respectivos autores para descrever os assuntos principais que permeiam o trabalho de forma clara e objetiva, sem se preocupar com questões semânticas (MCCLOSKEY, 1998; KHAN; WOOD, 2015). Com isso, neste trabalho, palavras-chave de artigos científicos são igualmente referenciadas como tópicos de pesquisas.

Dentre os trabalhos encontrados, geralmente as análises realizadas utilizam repositórios internacionais, que são específicas de uma determinada área ou periódico. No entanto, por tratarem de análises específicas e utilizarem de repositórios internacionais, não podem representar de forma abrangente o que é produzido no Brasil. Com isso, analisar fonte de dados que englobe diversos tipos de publicações, principalmente em veículos nacionais de diversas áreas, passa a ser uma tarefa relevante para a compreensão da ciência brasileira.

Neste contexto, Ferraz et. al. (2014) destacam que o repositório da Plataforma Lattes representa o instrumento mais importante para estudos sobre a produção científica brasileira. Contudo, mesmo estando disponível livremente na internet, esses dados ainda não foram

amplamente analisados (DIGIAMPIETRI, 2015). Dos trabalhos que exploram estes dados, poucos são os que analisam as palavras-chave das publicações. Geralmente utilizam termos extraídos dos títulos dos artigos de um conjunto restrito de currículos, na tentativa de destacar os principais assuntos abordados.

Assim sendo, este trabalho tem como objetivo realizar uma análise temporal sobre as palavras-chave de artigos publicados entre 1997 e 2016 em periódicos pelos doutores que possuem currículos cadastrados na Plataforma Lattes e, dedicaram esforços à grande área Ciência Sociais Aplicadas. O estudo contempla análises bibliométricas, afim de identificar e mapear os principais tópicos de interesses dos doutores brasileiros que estão atuando ou que já atuaram na área de Ciências Sociais Aplicadas, evidenciando-os que possuem destaque em determinada época tanto do ponto de vista quantitativo (frequência de aparecimento) quanto qualitativo (relevância em que o artigo que contempla determinado tópico foi publicado).

2 TRABALHOS CORRELATOS

Esforços para identificação e análises de tópicos de pesquisa constituem uma forma de melhorar a compreensão do que se tem produzido acerca da ciência e, a partir disto, auxiliar nos mais variados cenários quanto à tomada de decisão. De acordo com Choi et. al. (2011), estes estudos podem ser baseados tanto em termos extraídos dos títulos das publicações quanto das palavras-chave de produções bibliográficas. Por exemplo, Souza et. al. (2015) analisaram uma rede de termos extraídos dos títulos de 11.303 artigos da área de Sistemas Embarcados cadastrados na *IEEEXplorer*. Após destacarem os principais tópicos, realizaram uma análise temporal e aplicaram regressão linear para gerar a equação de previsão de tendências de pesquisas nos próximos cinco anos.

O trabalho de Liu et. al. (2012) consistiu em analisar 9.538 palavras-chave dos 2.647 artigos das principais revistas do repositório *Full Text*, para mapear a estrutura intelectual da área de “biblioteca digital” desenvolvida na China. Assim, foi construída uma matriz de co-ocorrência com base nas palavras-chave mais frequentes, para auxiliar na construção do agrupamento hierárquico dos tópicos e da rede para análises.

Pollack e Adler (2015) aplicaram análises bibliométricas para descobrir as tendências de pesquisas da área de Gerenciamento de Projetos a partir de 94.472 artigos da *Scopus* e *Web of Science* publicados entre 1962 e 2012. Eles realizaram comparações entre as palavras-chave e

os termos de resumos dos artigos quanto à frequência, sua variação de utilização ao longo do tempo e a co-ocorrência para identificar os tópicos emergentes.

Zhu et. al. (2015) analisaram as palavras-chave de 363.458 teses publicadas entre 1986 e 2014 referente a indústria de petróleo e gás da *China National Knowledge Internet*, para entender o desenvolvimento da área e evidenciar os assuntos atuais. Inicialmente, realizaram um processo para mineração dos dados, e, em seguida, analisaram as frequências, co-ocorrências e as redes das palavras-chave.

Medeiros e Mena-Chalco (2013) analisaram mais de 650 mil currículos cadastrados na Plataforma Lattes a fim de estudar a rede social dos indivíduos que atuam em áreas distintas. Adicionalmente às redes, analisaram as frequências dos termos extraídos dos títulos para identificar quais são mais utilizados. Para tanto, utilizaram mapas de termos e árvores de mapas para as 200 palavras mais frequentes em cada área.

O trabalho de Digiampietri et. al. (2014) apresentou uma análise sobre a rede social referente a 2002 currículos de indivíduos que possuem cadastros na Plataforma Lattes e que atuam no Brasil em Inteligência Artificial. Adicionalmente às redes, analisaram a frequência dos termos contidos nos títulos dos artigos publicados entre 1993 a 2012 para mostrar a evolução dos temas ao longo de cada ano.

Trucolo e Digiampietri (2014a) aplicaram regressões nos índices de importância *TF-IDF* dos termos extraídos de 34.289 títulos de artigos publicados pelos doutores da Ciência da Informação entre 1991 e 2012, para identificar tendências. Assim, um termo foi considerado tendência se possuíse uma alta previsão de *TF-IDF*. Trucolo e Digiampietri (2014b) aplicaram as mesmas regressões nos termos de 57.501 títulos de artigos publicados entre 1911 e 2011 pelos professores de Ciência da Computação.

Diante do exposto, nota-se que a identificação dos principais temas de pesquisa é viabilizada através da utilização das medidas de importância de tópicos com base na frequência dos termos ou em medidas de centralidade de redes desses termos. Apesar da validade dos resultados apresentados, no caso da Plataforma Lattes, que possui dados referentes a artigos publicados em periódicos e conferências de diferentes níveis de impacto, análises que contemplam apenas atributos quantitativos não conseguem explorar em completude as características qualitativas existentes nestes dados. À vista disso, o presente trabalho apresenta-se como a análise quanti-qualitativa sobre as palavras-chave dos artigos publicados pelos

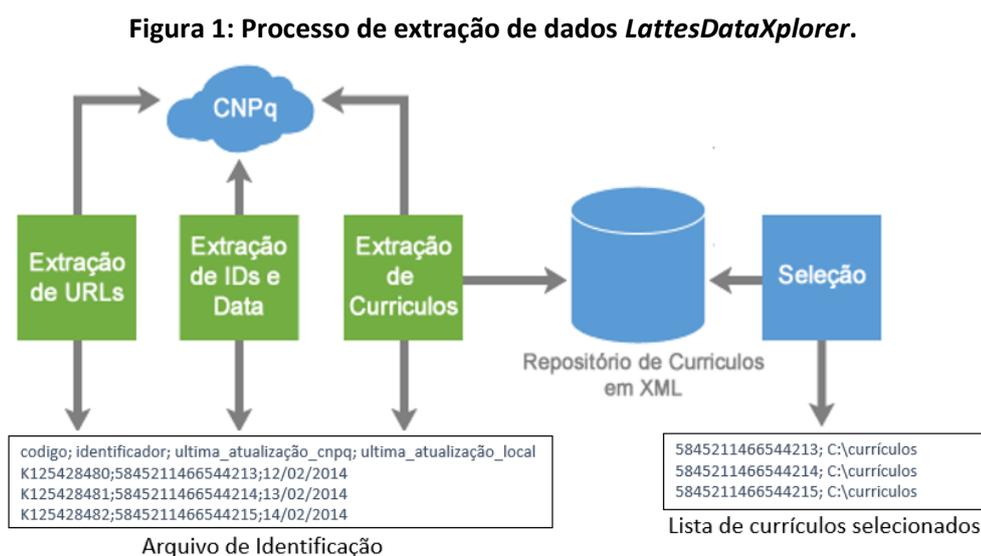
doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes e, dedicaram esforços a grande área Ciências Sociais Aplicadas.

3 DESENVOLVIMENTO

A escolha da Plataforma Lattes como fonte de dados está relacionada: (1) aos dados estarem disponíveis na internet e não terem sido amplamente analisados (DIGIAMPIETRI, 2015); (2) tratar da integração de dados de produções científicas de todas as áreas de C&T existentes na ciência brasileira; (3) por não negligenciar os artigos publicados em periódicos nacionais que muitas vezes não são indexados, e também os artigos de anais de congresso (DIAS, 2016); e, (4) por ser uma poderosa fonte para fornecimento de dados de alta qualidade para medir e avaliar o desempenho acadêmico nacional (LANE, 2010).

Neste ponto, vale ressaltar que, apesar dos dados disponibilizados, estes são apenas visualizados através de uma interface de consulta que apresenta apenas os dados de um único currículo. Logo, para uma análise mais detalhada de grupos, instituições ou até toda a nação de cientistas brasileiros, técnicas e ferramentas para análises dos dados se fazem necessárias.

A aquisição dos currículos dos doutores na versão XML foi realizada através da utilização do *LattesDataExplorer*, desenvolvido por Dias (2016) para coletar os dados científicos contidos repositório curricular da Plataforma Lattes (Figura 1).

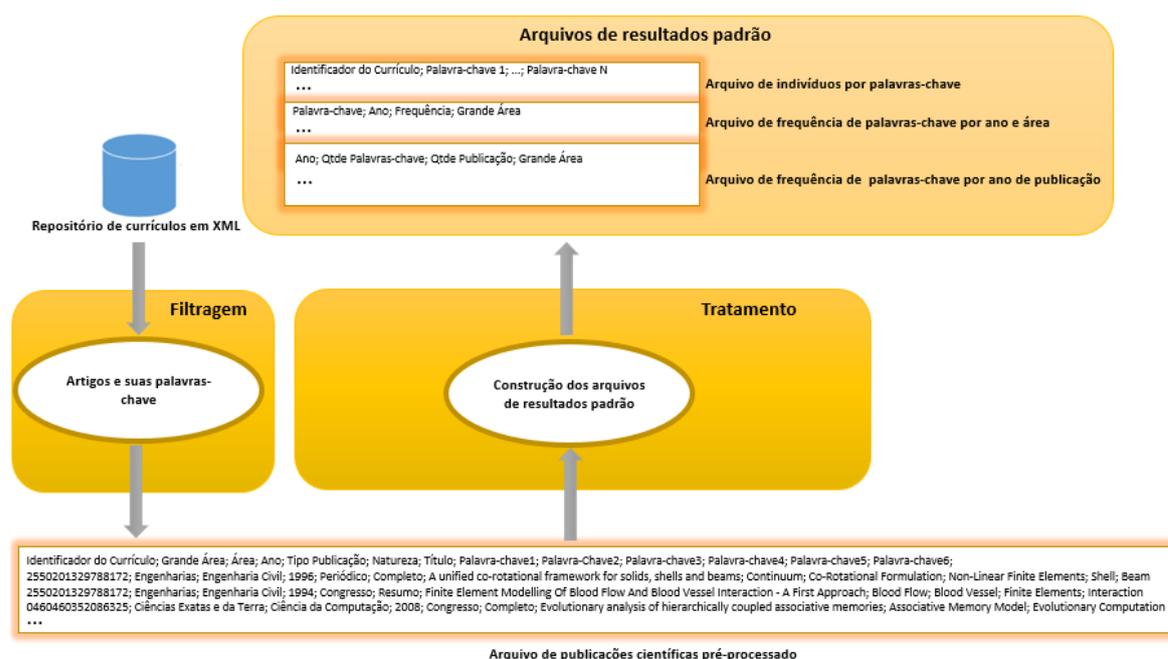


Fonte: Dias (2016)

A extração dos dados foi realizada em três etapas: (1) extração de URLs, responsável por adquirir as referências únicas para os currículos cadastrados, (2) extração de Ids, para possibilitar o acesso a cada currículo e extrair seu identificador, e (3) extração dos currículos, responsável pelo *download* e armazenamento em disco dos arquivos (DIAS, 2016).

A Figura 2 apresenta uma visão detalhada do arcabouço de componentes desenvolvido para suportar as análises desejadas neste trabalho. Nela, os componentes de “filtragem dos dados” e “tratamento dos dados” são os responsáveis pelo processo de seleção, tratamento e caracterização dos dados dos currículos que realmente necessitam ser processadas, e, concomitantemente, diminuir o tempo de processamento computacional.

Figura 2: Arcabouço de componentes desenvolvidos para filtragem e tratamento dos dados.



Fonte: Elaborado pelos autores

De posse dos currículos XML, o componente de “filtragem dos dados” realiza a etapa de mineração nos currículos para extrair as informações dos artigos, armazenando-as à parte num arquivo de publicações científicas, e com isso, definindo o conjunto de dados centrais a serem estudados. Esses dados incluem: identificador do currículo; grande área da publicação; ano de publicação; idioma, tipo de publicação; título e palavras-chave.

Em contrapartida, o componente de “tratamento dos dados”, processa os dados do arquivo de publicações científicas para tratá-los e caracterizá-los, e, a partir destes construir um conjunto dos arquivos de resultados padrão para facilitar as análises. Esse componente realiza

basicamente três etapas, a saber: *limpeza e agrupamento dos dados, identificação da colaboração científica e construção dos arquivos*.

A *limpeza dos dados* serve para realizar o processamento das palavras-chave de tal forma a excluir os termos que não representam tópicos de estudos e agrupar as palavras que possuem mesmo valor semântico. Para isso, inicialmente, o método desenvolvido obtém as palavras-chave de cada artigo analisado. Em seguida, cada uma das palavras-chave são associadas ao idioma cadastrado para o artigo, servindo de referência no processo de radicalização. O processo de lowercase converte todas as palavras para minúsculo no intuito de padronizar o conjunto. No processo de stopWords são removidos os termos que não possuem valores semânticos. Posteriormente, no processo de normalização todos os acentos e pontuações são retirados das palavras-chave. Finalmente, o processo de radicalização é responsável pela redução da palavra-chave a seu radical. Contudo, em caso de palavras-chave compostas, este processo é executado em cada termo individualmente, e, concatenado formando uma única palavra.

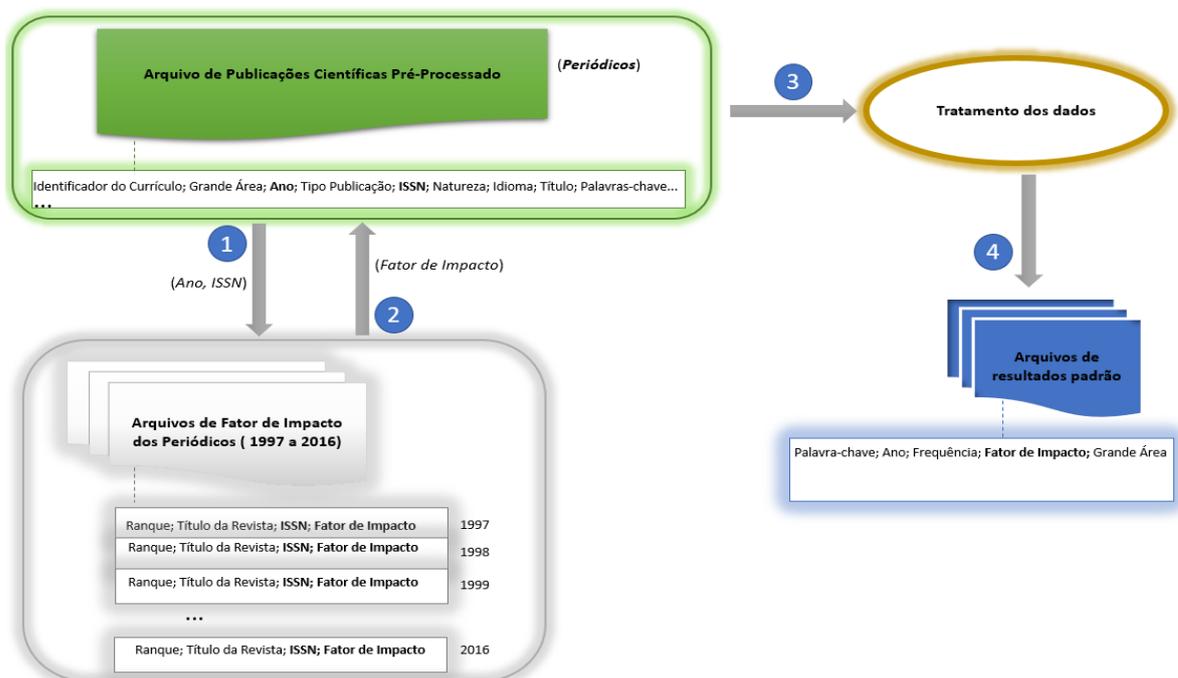
Na etapa de *identificação de colaboração científica*, todos os artigos do arquivo de publicações científicas são processados para identificar e tratar os trabalhos realizados em colaboração. Esse procedimento é realizado para dados de um mesmo artigo científico não ser considerado mais de uma vez durante as análises. Para tanto, fez-se necessário uma adaptação no método ISCooll proposto por Dias e Moita (2015) para identificar colaborações científicas em grande volume de dados. Enquanto o ISCooll original utiliza um dicionário para vincular os artigos a seus autores, o ISCooll adaptado adota um dicionário para vincular os artigos ao conjunto união das palavras-chave. Outra diferença está na formação da chave do dicionário; enquanto ISCooll original utiliza o título do artigo concatenado com o ano de publicação, o ISCooll adaptado adiciona também a grande área da publicação.

A construção *dos arquivos* serve para facilitar a extração de conhecimento durante as análises e contribuir na diminuição dos dados a serem processados. Nesta etapa, foi construído um conjunto de arquivos de resultados padrão para cada tipo de publicação. Os arquivos de resultados padrão formam o repositório principal que possibilita as análises estatísticas sobre os tópicos de pesquisas desenvolvidos. Inicialmente, foram construídos três arquivos de seguintes formatos: (1) *arquivo de indivíduos por palavras-chave* (identificador do currículo, palavra-chave1,... palavra-chaveN); (2) *arquivo de palavras-chave por ano e grande área* (palavra-chave, ano, frequência e grande área da publicação) e (3) *arquivo de frequência de*

palavras-chave por ano e quantidade de publicação (ano, frequência palavra-chave, quantidade de publicação e grande área da publicação).

Em continuação, tem-se que periódico é um dos meios de comunicação formal mais utilizado pelos pesquisadores para disseminação dos resultados de suas pesquisas na maioria das áreas do conhecimento. Com base nisso, foram selecionados os dados dos artigos publicados em periódicos pelos doutores brasileiros da grande área Ciências Sociais Aplicadas contido no arquivo de publicações científicas, para representar o conjunto das informações a serem exploradas. Contudo, sabe-se que a publicação de artigos em periódicos de prestígio possibilita ao pesquisador reconhecimento junto a seus pares, citações por outros autores e visibilidade no cenário nacional e/ou internacional, por isso é importante considerar a qualidade e confiabilidade de cada periódico para tornar as análises mais coesas ao contexto.

Assim sendo, os dados referentes aos artigos dos doutores brasileiros contidos no arquivo de publicações científicas foram complementados com o valor do Fator de Impacto (FI) do respectivo periódico. Em seguida, o componente de “tratamento dos dados” foi ajustado para adicionar a somatória dos valores de *FI* dos periódicos por ano às respectivas palavras-chave dos artigos contidas nos arquivos de resultado padrão. Dessa maneira, as palavras-chave de artigos publicados em periódicos de relevância tem valor maior de FI em relação as palavras-chave de artigos publicados em periódicos de menor impacto. Isso foi possível através da utilização do conjunto de dados do JCR (*Journal Citation Reports*), pois, de posse do ISSN do periódico, permite obter informações referentes ao impacto e influência deste periódico na comunidade científica ano a ano desde 1997. A Figura 3 ilustra a adaptação necessária no arcabouço de componentes para a inserção do valor de *FI* dos periódicos no conjunto dos dados contidos no arquivo de publicações científicas e nos arquivos de resultado padrão.



Fonte: Elaborado pelos autores

Dessa forma, o conjunto de dados de publicações em periódicos selecionado para esse estudo passa a contemplar o Fator de Impacto destes e refere-se ao período de 1997 a 2016, ou seja, data inicial de disponibilização do *FI* pela JCR ao último ano considerado nas análises. Neste ponto, é importante destacar que os periódicos cadastrados nos currículos dos doutores brasileiros e não indexados em bases internacionais receberam o valor 0 de *FI*.

Para possibilitar levar em consideração tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados, foi desenvolvida a medida de importância de tópico *TF.FI*. Nesse caso, a *TF.FI* pondera a frequência das palavras-chave de um conjunto de artigos publicados em periódicos com a soma dos valores de *FI* dos periódicos em que foram publicados os artigos que contém as respectivas palavras-chave. Essa medida parte do pressuposto que palavras-chave de um artigo publicado em um periódico de alta relevância são mais importantes que as mesmas palavras-chave de um outro artigo publicado em periódico de menor relevância, e, por isso, as palavras-chave de artigos publicados em periódicos de destaque recebem maior valor de *FI* como parcela de ponderação.

A Figura 4 apresenta a equação para o cálculo da medida *TF.FI* de um determinado tópico num intervalo de tempo, para posteriormente esses tópicos serem ranqueados no intuito de destacar os mais importantes do conjunto de dados em análise.

Figura 4: - Medida de importância de tópicos TF.FI.

$$TF.FI_{K(i,m)} = \sum_{n=i}^m TF_{K(n)} \cdot \sum_{n=i}^m FI_{K(n)}$$

Onde:

$TF_{K(n)}$ = Quantidade da palavra-chave k inserida nos artigos publicados em determinado ano

$FI_{K(n)}$ = Soma dos valores de FI dos periódicos onde foram publicados os artigos que contém a palavra-chave k em determinado ano

i = Ano inicial do período de análise

m = Ano final do período de análise

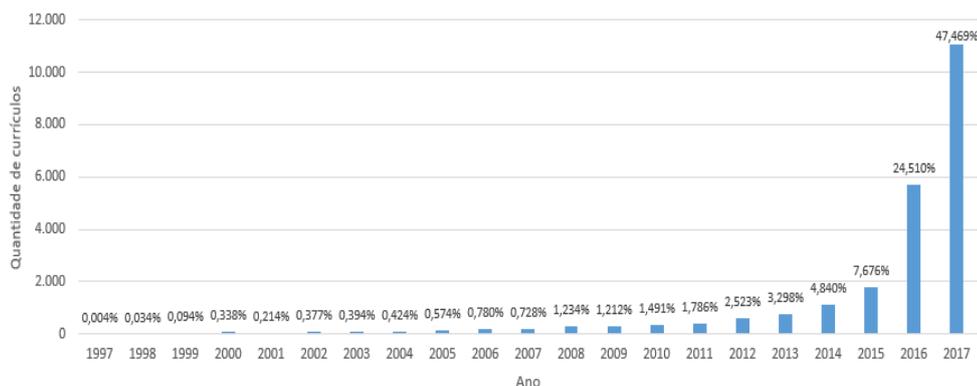
Fonte: Elaborado pelos autores

Como pode ser observado na figura, para um determinado tópico de pesquisa possuir destaque a partir da aplicação da medida de importância TF.FI em um conjunto de dados, é necessário que este tópico tenha um considerável índice de frequência e que também esteja inserido em artigos publicados por periódicos com maior valor de fator de impacto em relação ao conjunto dos dados analisados.

4 RESULTADOS

Os dados foram coletados em abril de 2017, através da etapa de *aquisição dos dados* descrito na Seção de “Desenvolvimento”, totalizando 265.170 currículos de indivíduos com doutorado concluído, onde 26.508 destes currículos têm as Ciências Sociais Aplicadas como principal grande área de atuação. Para visualizar a contemporaneidade do conjunto de dados a serem analisados, a Figura 5 apresenta a distribuição dos currículos dos doutores que atuam nas Ciências Sociais Aplicadas com base na última data de atualização.

Figura 5: Distribuição dos currículos dos doutores pela data da última atualização.



Fonte: Elaborado pelos autores.

Como pode ser observado, a grande maioria dos currículos foram atualizados recentemente, onde aproximadamente 48% possuem data de última atualização em 2017 e cerca de 72% atualizados nos últimos dois anos. A não atualização de parte desses currículos pode ter motivos variados e de difícil reconhecimento (DIAS, 2016).

Para as análises foram considerados todos os artigos publicados pelos doutores em anais em periódicos entre 1997 e 2016 referentes a grande área Ciências Sociais Aplicadas. Entretanto, cada indivíduo possui apenas um currículo cadastrado na Plataforma Lattes, não existindo a possibilidade de dados duplicados sobre a produção deste. Porém, se considerado os dados de todos os currículos, um artigo produzido em colaboração pode aparecer várias vezes, tendo em vista que ele pode ser registrado por cada um de seus autores.

Neste ponto, com o propósito de tornar as análises mais coesas ao contexto, foi necessária a consideração de duas visões sobre os dados extraídos: **a visão geral**, que contabiliza todos os artigos e palavras-chave extraídos dos currículos e, **a visão colaboração**, que, considera o artigo publicado em coautoria de mesma grande área uma única vez associado ao conjunto união de suas palavras-chave. Sendo assim, a Tabela 1 apresenta o quantitativo dos dados de artigos extraídos dos currículos dos doutores que contribuíram com a grande área Ciências Sociais Aplicadas com data de publicação entre 1997 e 2016.

Tabela 1: Quantitativo de artigos científicos e suas palavras-chave dos currículos dos doutores.

	Tipo de Publicação	Artigos	Palavras-chave vinculadas aos artigos
Geral	Periódicos	343.270	603.067
Colaboração	Periódicos	291.937	609.524

Fonte: Elaborado pelos autores.

Diante disso, percebe-se que o número de artigos na visão colaboração é inferior ao respectivo número na visão geral, isso acontece devido artigos realizados em coautoria serem considerados apenas uma única vez. Já quanto o número de palavras-chave na visão colaboração ser maior que a visão geral, é devido os doutores terem associados diferentes grandes áreas e diferentes palavras-chave para o mesmo artigo publicado em conjunto. Outra situação está relacionada à não obrigatoriedade de inserção das palavras-chave aos artigos pela Plataforma Lattes, o que implica em cerca de 45% dos artigos na visão geral e 36% na colaboração não possuir nenhuma palavra-chave associada.

Entretanto, para as análises seguintes será considerado apenas o conjunto de dados referente a visão colaboração. Tal escolha deve-se a basicamente a dois motivos: (1) devido

um mesmo trabalho realizado em coautoria não conter palavras-chave repetidas, pois, caso contrário, o cálculo da frequência destas palavras poderia ser impactado pela quantidade de dados inseridos por seus autores, (2) serem consideradas todas as diferentes palavras-chave inseridas pelos respectivos autores para o mesmo trabalho que realizaram conjuntamente, aumentando assim a possibilidade de melhorar a descrição dos assuntos principais que permeiam o artigo.

Em continuação, o conjunto global (609.524) das palavras-chave referente a visão colaboração foram processados conforme descrito na seção anterior, e, se reduziu a 125.506 palavras-chave únicas, comprovando a importância do tratamento dos dados para a diminuição das informações a serem processadas. Em seguida, para facilitar a compreensão da evolução temporal dos principais temas abordados pelos doutores brasileiros que atuam nas Ciências Sociais Aplicadas ao longo do tempo, todas as palavras-chave únicas foram ranqueadas de acordo com a medida de importância de popularidade TF.FI a cada período de 2 anos entre 1997 e 2016. Assim, a Tabela 2 apresenta as principais palavras-chave de acordo com sua frequência de utilização em cada biênio entre todo o histórico analisado.

Tabela 2: As principais palavras-chave das Ciências Sociais Aplicadas ranqueadas em cada biênio.

Ranque	1997-1998	TF.FI	1999-2000	TF.FI	2001-2002	TF.FI
1	Tecnologia da Informação	124	Sequenciamento	1.440	Brasil	770
2	Inflação	84	Genoma	1.260	Estratégia	273
3	Brasil	70	Bioinformática	231	Desenvolvimento	234
4	Bibliometria	55	Agente patogênico	231	Discurso	128
5	Marketing	38	Xylella Fastidiosa	231	Família	96
6	Amazônia	18	DNA	204	Economia Política	62
7	Sindicato	15	Políticas Públicas	146	Tecnologia da Informação	55
8	Organização Industrial	12	Genética	125	Base de Dados	52
9	Cientometria	9	Laranja	125	Energia	46
10	Microeconometria	9	Qualidade de Vida	110	Qualidade de Vida no Trabalho	44
11	Firma	9	Citrus	102	Indústria	34
12	Infometria	6	CVC	102	Sustentabilidade	31
13	Ciência da Informação	6	Fitopatôgeno	102	Pobreza	29
14	Gestão do Conhecimento	4	Planejamento Urbano	56	Amazônia	28
15	Web 2.0	3	Amazônia	36	Mulher	28
Total de Palavras-chave do Biênio		20.124		27.444		35.847
Ranque	2003-2004	TF.FI	2005-2006	TF.FI	2007-2008	TF.FI
1	Amazônia	1.672	Sustentabilidade	735	Gênero	1.960
2	Brasil	708	Comunicação	405	Sustentabilidade	1.810
3	Sustentabilidade	585	Saúde	396	Brasil	930
4	Cultura	132	Brasil	350	Mercado de Trabalho	693
5	Mercado de Trabalho	108	Competitividade	306	Estratégia	600
6	Saúde	92	Qualidade de Vida	300	América Latina	497
7	Economia	75	Amazônia	270	Produção Científica	448
8	Governança	72	Empreendedorismo	232	Design	392
9	Capitalismo	68	Antropologia	208	Amazônia	325
10	Crescimento Econômico	62	Reforma Agrária	192	Pobreza	305
11	Peptídeos	48	Educação	164	Governança	261
12	Desenvolvimento Rural	40	Produção Científica	150	Educação	196
13	Pobreza	37	Financiamento	148	Marketing	187
14	Bioética	36	Planejamento Urbano	138	Inovação	165
15	Comércio Eletrônico	36	Avaliação	132	Gestão do Conhecimento	159
Total de Palavras-chave do Biênio		46.639		56.636		63.855
Ranque	2009-2010	TF.FI	2011-2012	TF.FI	2013-2014	TF.FI
1	Inovação	2.460	Brasil	6.688	Marketing	5.962
2	Brasil	2.430	Amazônia	2.700	Inovação	5.180

XIX ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2018
22 a 26 de outubro de 2018 – Londrina – PR

3	Saúde	2.176	Inovação	2.604	Brasil	4.914
4	Consumismo	1.581	Design	2.184	Bibliometria	3.325
5	Amazônia	1.224	Sustentabilidade	1.820	Amazônia	3.036
6	Marketing	955	Gestão do Conhecimento	1.323	Sustentabilidade	1.980
7	Gestão do Conhecimento	796	Agronegócio	1.253	Epidemiologia	1.584
8	Educação	786	Marketing	1.225	Agricultura Familiar	1.467
9	Políticas Públicas	748	Comunicação	1.122	Administração	1.160
10	Design	650	Estratégia	1.092	Estratégia	1.146
11	Turismo	645	Biodiesel	1.080	Gestão do Conhecimento	930
12	Internet	588	Políticas Públicas	1.004	Agricultura	910
13	Ciência da Informação	417	Governo	925	Gênero	852
14	Bibliometria	414	Consumismo	903	Cultura	776
15	Energia	396	Energia	888	Gestão Pública	690
Total de Palavras-chave do Biênio		77.286		89.261		96.237
Ranque	2015-2016	TF.FI	1997-2016	TF.FI		
1	Brasil	9.100	Brasil	176.349		
2	Comunicação	4.491	Amazônia	98.464		
3	Sustentabilidade	4.096	Marketing	84.182		
4	Bibliometria	3.380	Inovação	79.648		
5	Inovação	3.090	Comunicação	54.950		
6	Universidade	2.720	Sustentabilidade	49.140		
7	Educação	2.601	Bibliometria	39.900		
8	Marketing	2.574	Educação	35.460		
9	Amazônia	2.254	Estratégia	33.840		
10	Interdisciplinar	1.716	Saúde	31.450		
11	Empreendedorismo	1.416	Design	29.512		
12	Estratégia	957	Políticas Públicas	28.960		
13	Agricultura Familiar	912	Gênero	27.864		
14	Internet	798	Gestão do Conhecimento	26.841		
15	Ensino Superior	756	Consumismo	21.917		
Total de Palavras-chave do Biênio		96.195		609.524		

Fonte: Elaborado pelos autores

Notadamente, a palavra-chave “Brasil” (tema de investigação genérico) tem destaque na grande área Ciências Sociais Aplicadas ao longo do tempo, visto seu alto valor de popularidade e devido ter figurado entre as principais palavras-chave em praticamente todos os biênios analisados, exceto no período de 1999-2000. Isso indica que historicamente os doutores que atuam na grande área Ciências Sociais Aplicadas têm dedicado esforços em pesquisas de interesses do país. As palavras-chave “Amazônia”, “Sustentabilidade” e “Consumismo”, possivelmente relacionadas a temas que envolvem problemas ambientais, contaram também com uma atenção especial por partes dos doutores, principalmente, “Amazônia” que figurou-se entre as principais em todos os biênios analisados.

A área que abrange temas relacionados a Comunicação, Marketing e Vendas também recebe grande atenção, como pode ser notado, as palavras-chave “Marketing”, “Comunicação”, “Inovação”, “Estratégia” e “Design” estão entre as mais populares considerando todo o período de análise dos dados. Sendo que a utilização destas palavras-chave pode estar relacionada com a modernização que o Marketing vem tendo desde a década de 90 com a evolução e popularização da internet, redes sociais e dispositivos móveis. Também vale destacar temas que podem estar relacionado a mais de uma linha de estudo, como é o caso da palavra-chave “Gênero”, que pode ser referente a estudos que tentam

compreender as relações de gêneros (feminino, masculino e transgeneridade) na cultura e sociedade humana ou gêneros gramaticais.

As palavras-chave “Saúde”, “Educação” e “Políticas Públicas” são considerados temas centrais para desenvolvimento de qualquer nação, e por isso, também receberam grande dedicação dos doutores que atuam na grande área Ciências Sociais Aplicadas. Em contrapartida, a popularidade das palavras-chave “Bibliometria” e “Gestão do Conhecimento” se deve ao fato certamente da Ciência da Informação ser uma das áreas da grande área Ciências Sociais Aplicadas, ou pelo menos era até pouco tempo atrás e por isso identificada nos currículos. Outra informação interessante é que a evolução do quantitativo das palavras-chave em cada biênio, isso devido ao aumento de esforços por parte dos doutores que atuam nas Ciências Sociais Aplicadas nos respectivos temas, que, possivelmente seguiu o mesmo comportamento do crescimento da produção científica brasileira.

Adicionalmente, na tentativa de mapear a evolução dos principais interesses científicos por parte dos doutores que atuam na grande área Ciências Sociais Aplicadas, a Tabela 3 apresenta o coeficiente de similaridade das 15 palavras-chave mais populares em cada par de biênios do período analisado. Neste contexto, coeficiente de similaridade mede o quão dois conjuntos de palavras-chave são iguais (a taxa da interseção entre eles), independentemente da ordem do ranqueamento destas palavras-chave em cada conjunto.

Tabela 3: Coeficiente de similaridade das principais palavras-chave das Ciências Sociais Aplicadas ranqueadas por TF.FI entre os biênios analisados.

	1997-1998	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	2015-2016
1997-1998	X	6,66%	20,00%	13,33%	13,33%	20,00%	26,66%	20,00%	26,66%	26,66%
1999-2000		X	6,66%	6,66%	20,00%	6,66%	13,33%	13,33%	6,66%	6,66%
2001-2002			X	26,66%	20,00%	33,33%	20,00%	33,33%	26,66%	26,66%
2003-2004				X	26,66%	40,00%	20,00%	20,00%	26,66%	20,00%
2005-2006					X	33,33%	26,66%	26,66%	20,00%	40,00%
2007-2008						X	46,66%	53,33%	53,33%	46,66%
2009-2010							X	60,00%	40,00%	46,66%
2011-2012								X	46,66%	46,66%
2013-2014									X	53,33%
2015-2016										X

Fonte: Elaborado pelos autores

Como pode ser notado, não houve ao longo do período analisado 100% de similaridade dos tópicos centrais estudados entre pares de biênios. Isso mostra que a cada 2 anos os principais interesses científicos dos doutores que atuam na grande área Ciências Sociais Aplicadas tem mudado por algum motivo. No entanto, ao comparar os pares de biênios

sequentes, nota-se que em 100% das comparações realizadas revelam que parte dos principais tópicos de pesquisas estudados no biênio atual foram considerados relevantes no biênio anterior, destacando os períodos de 2009-2010 e 2011-2012, onde o percentual de similaridade foi de 60,00%, divergindo quanto as palavras-chave “Biodiesel”, “Comunicação”, “Governo”, “Agronegócio”, “Estratégia” e “Sustentabilidade” que tiveram popularidade destacada em 2011-2012; enquanto que, “Ciência da Informação”, “Bibliometria”, “Turismo”, “Internet”, “Educação” e “Saúde” foram destaque em 2009-2010.

Outra possibilidade é a análise do coeficiente de similaridade dos tópicos de pesquisas entre os biênios não sequentes, pois é permitido verificar se tópicos que foram destaques num passado mais distante mantiveram, desapareceram ou retornaram em evidência em períodos mais recentes. Neste contexto, destaca-se que somente a palavra-chave “Amazônia” evidenciada nos biênios de 2013-2014 e 2015-2016 também apareceu entre os principais tópicos estudados no período de 1999-2000.

5 CONSIDERAÇÕES FINAIS

Este estudo contribuiu para a identificação dos tópicos de pesquisas que obtiveram maior interesses em cada época por parte dos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes e atuam na grande área Ciências Sociais Aplicadas. Ao realizar uma análise temporal entre as principais palavras-chave de cada período, foi possível verificar que as preferências centrais dos doutores sobre os tópicos de pesquisas foram alterando gradualmente ao longo do tempo, possivelmente devido a uma demanda externa da sociedade ou por determinado assunto ter atingido maturidade.

Tais análises caracterizam-se como um importante mecanismo, pois podem identificar quais tópicos mais impactantes são desenvolvidos no Brasil e, assim, servir como apoio a diversos tipos de tomada de decisão. Como contribuição à validade dos resultados, é possível destacar a relevância do repositório da Plataforma Lattes para estudos bibliométricos sobre o desenvolvimento da ciência brasileira, aliado a abrangência e contemporaneidade de atualização dos dados considerados para análises.

A sistemática apresentada para a filtragem, tratamento e manipulação das palavras-chave contidas nos artigos de currículos da Plataforma Lattes forneceu uma metodologia nova para estudos sobre a identificação e a análises de tópicos de pesquisas. O mesmo procedimento pode ser replicado em outros conjuntos de palavras-chave de artigos científicos.

Como alternativa às medidas de importância de tópicos com base na frequência e grau que consideram apenas atributos quantitativos, foi proposto uma medida de importância de tópico (TF.FI) que é capaz de levar em consideração tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados. Assim, a medida de TF.FI consegue explorar em completude a heterogeneidade de níveis de impacto dos periódicos cadastrados na base curricular da Plataforma Lattes.

Com isso, como trabalhos futuros, espera-se que sejam realizadas análises que consideram fatores temporais para determinar a relevância de um tópico em cada época, e, análises baseadas em técnicas de redes sociais para determinar os relacionamentos entre as palavras-chave. Além disso, devido à pouca padronização existente no registro das palavras-chave por parte dos pesquisadores quanto as áreas do conhecimento, espera-se também a utilização de um thesaurus para tal finalidade.

AGRADECIMENTOS

Os autores agradecem ao CEFET-MG pelo auxílio na pesquisa.

REFERÊNCIAS

- BORGES, V. A. et. al. Uma análise exploratória de tópicos emergentes em Informática na Educação. **Revista Brasileira de Informática na Educação**, v. 23, n. 1, p. 1-13, 2015. Disponível em: <<http://dx.doi.org/10.5753/rbie.2015.23.01.85>> Acesso em: 10 jan. 2017.
- BRITO, A. G. C. et. al. Exploração da Plataforma Lattes por assunto: proposta de metodologia. **Transinformação**, v.28, n.1, p. 77-86, 2016. Acesso em: 03 jan. 2017. Disponível em: <<http://periodicos.puc-campinas.edu.br/seer/index.php/transinfo/article/view/2519>>.
- CASTELLS, M. A sociedade em rede. **Paz e Terra**, 2010.
- CAVACINI, A. Recent trends in Middle Eastern scientific production. **Scientometrics**, v. 109, n. 1, p. 423-432, 2016. Disponível em: <<http://link.springer.com/article/10.1007/s11192-016-1932-3>>. Acesso em: 03 jan. 2017.
- CHEN, G. et. al. Identifying the research focus of Library and Information Science institutions in China with institutions-specific keywords. **Scientometrics**, v. 103, n. 2, p. 707-724, 2015. Disponível em: <<http://link.springer.com/article/10.1007/s11192-015-1545-2>>. Acesso em: 05 jan. 2017.
- CHOI, J. et. al. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. **Information & Management**, v. 48, n. 8, p. 371-381, 2011. Disponível em: < <http://dx.doi.org/10.1016/j.im.2011.09.004> > Acesso em: 15 jan. 2017.

- DIAS, T. M. R. **Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes**. 2016. 181 (Doutorado). Pós-Graduação em Modelagem Matemática e Computacional, PPGMMC/CEFET-MG, 2016.
- DIAS, T. M. R.; MOITA, G. F. A method for the identification of collaboration in large scientific databases. **Em Questão**. v.21, n.2, 2015. Acesso em: 29 jan. 2017. Disponível em: <<http://seer.ufrgs.br/index.php/EmQuestao/article/view/53259>>
- DIGIAMPIETRI, L. A. **Análise da rede social acadêmica brasileira**. 2015. 160. (Livre Docência). Escola de Artes, Ciências e Humanidades da Universidade de São Paulo USP, São Paulo, 2015.
- DIGIAMPIETRI, L. A.; PERES, S. M.; SILVA, L. A. Rede de Relacionamentos Brasileira de Inteligência Artificial e Computacional. **XI Encontro Nacional de Inteligência Artificial e Computacional (ENIAC 2014)**, p. 1-6, 2014, São Carlos, SP, Brasil.
- FERRAZ, R. R. N. et. al. The use of ScriptLattes tool for extraction and on-line availability of academic production from a department of stricto sensu in management. In: **International Conference on Information Systems and Technology Management**, SP, p. 663 – 679, 2014.
- FERREIRA, A. G. C. Bibliometria na avaliação de periódicos científicos. *Data Grama Zero*. **Revista de Ciência da Informação**, v. 11, n. 3, p. 1-9, 2010.
- KHAN, G. F.; WOOD, J. Information technology management domain: emerging themes and keyword analysis. **Scientometrics**, v. 105, n. 2, p. 959-972, 2015. Acesso em: 10 jan. 2017. Disponível em: <<http://link.springer.com/article/10.1007/s11192-015-1712-5>>
- LANE, J. Let's make science metrics more scientific. **Nature**, v. 464, n. 7288, p. 488-489, 2010.
- LIU, G. Y. et. al. A co-word analysis of digital library field in China. **Scientometrics**, v. 91, n. 1, p. 203-217, 2012. Disponível em: <<https://link.springer.com/article/10.1007/s11192-011-0586-4>> Acesso em: 25 jan. 2017.
- MEDEIROS, C.; MENA-CHALCO, J. P. The Dynamics of Multidisciplinary Research Networks - Mining a Public Repository of Scientists CVs. **World Social Science: social transformations and the digital ages**, 2013, Montreal, Canadá.
- MCCLOSKEY, D. N. (1998). **The rhetoric of economics**. University of Wisconsin Press.
- OHNIWA, R. L.; HIBINO, A.; TAKEYASU, K. Trends in research foci in life science fields over the last 30 years monitored by emerging trends. **Scientometrics**, v. 85, n. 1, p. 111-127, 2010. Disponível em: <<http://link.springer.com/article/10.1007/s11192-010-0252-2>>
- POLLACK, J.; ADLER, D. Emergent trends and passing fads in project management research: A scientometric analysis of changes in the field. **Project Management**, v. 33, n. 1, p. 236-248, 2015. Disponível em:<<http://dx.doi.org/10.1016/j.ijproman.2014.04.011>>.

RIDKER, P. M.; RIFAI, N. Expanding Options for Scientific Publication. Is More Always Better? *Circulation*, v. 127, n. 1, p.155-156, 2013.

SAES, S.G. **Aplicação de métodos bibliométricos e da co-word analysis na avaliação da literatura científica brasileira em ciências da saúde de 1990 a 2002**. 2005. (Doutorado). Pós-Graduação em Saúde Pública, USP, São Paulo, 2005.

SOUZA, J. et. al. Análise de rede de termos em Sistemas Embarcados através de análise da rede de termos em títulos de trabalhos científicos. **IV Brazilian Workshop on Social Network Analysis and Mining (BraSNAM 2015)**. Recife, Brasil, 2015. Disponível em: <<http://dx.doi.org/10.13140/RG.2.1.3246.3207>> Acesso em: 25 jan. 2017.

TRUCOLO, C. C.; DIGIAMPIETRI, L. A. Análise de tendências da produção científica nacional na área de Ciência da Informação: estudo exploratório de mineração de textos. **Atoz**: novas práticas em informação e conhecimento, v. 3, n. 2, p. 87-94, 2014a. Disponível em: <<http://revistas.ufpr.br/atoz/article/view/41341>> Acesso em: 25 jan. 2017.

TRUCOLO, C. C.; DIGIAMPIETRI, L. A. Trend Analysis of the Brazilian Scientific Production in Computer Science. **Revista de Sistemas de Informação da FSMA**. 2014b. p. 2–10.

ZHU L. et. al. Keywords co-occurrence mapping knowledge domain research base on the theory of Big Data in oil and gas industry. **Scientometrics**, v. 105, n. 1, p. 249-260, 2015. Disponível em: <<http://dx.doi.org/10.1007/s11192-013-1019-3>> Acesso em: 20 jan. 2017.