

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017

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

**IDENTIFICAÇÃO DAS PALAVRAS-CHAVE MAIS FREQUENTES REGISTRADAS NOS
CURRÍCULOS DA PLATAFORMA LATTES**

Jether Gomes (CEFET-MG)

Thiago M. R. Dias (CEFET-MG)

Gray F. Moita (CEFET-MG)

***IDENTIFICATION OF THE MOST FREQUENT KEYWORDS REGISTERED IN LATTES PLATFORM
CURRICULA***

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 é o grande número de informações disponibilizadas na internet e a ampliação da sociabilização entre os pesquisadores. Conseqüentemente, 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. 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 entre 1962 e 2016 em anais de congressos e em periódicos, que são ranqueadas temporalmente através da aplicação de duas medidas de popularidade de tópicos baseado em frequência. Os resultados possibilitaram uma análise temporal dos tópicos investigados pelos doutores a cada quinquênio e, também uma análise comparativa entre os conjuntos de palavras-chave ranqueados pelas medidas de importância de tópicos frequência e taxa de aparecimento. Como considerações finais, observa-se que os doutores brasileiros 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. Além disso, pode ser observado baixa

correlação entre os conjuntos de palavras-chave referentes as medidas de importância de tópicos empregadas.

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

Abstract: Science has presented 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 researchers who have curricula registered in the Lattes Platform. To do so, 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 article keywords published between 1962 and 2016 in congress proceedings and journals, which are ranked temporally through the application of two measures of popularity of topics based on frequency. The results allowed a temporal analysis of the topics investigated by the PhDs every five years and also a comparative analysis between the sets of keywords ranked by the measures of importance of topics frequency and rate of appearance. As final considerations, it is observed that the Brazilian PhDs devoted more effort to a reduced set of research topics, and that the exchange of major interests gradually evolved over time, possibly due to a society's demand or maturation of a subject. In addition, it can be observed the low correlation between the sets of keywords referring to the importance measures used.

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,

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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, nota-se que 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 de se preocupar com a semântica e estrutura dos termos.

Logo, uma abordagem interessante 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 1962 e 2016 em anais de congressos e em periódicos pelos doutores que possuem currículos cadastrados na Plataforma Lattes. O estudo contempla análises bibliométricas, afim de identificar e mapear os tópicos de interesses dos doutores brasileiros, evidenciando-os que possuem destaque em determinada época.

2 TRABALHOS CORRELATOS

Esforços para identificação e análises de tópicos de pesquisa podem auxiliar nos mais variados cenários quanto à tomada de decisão no contexto científico. 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 construído uma matriz de co-ocorrência com base nas palavras-chave mais frequentes, para auxiliar na a 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 possuísse 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, o presente trabalho apresenta-se como a análise abrangente sobre as palavras-chave dos artigos publicados pelos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes.

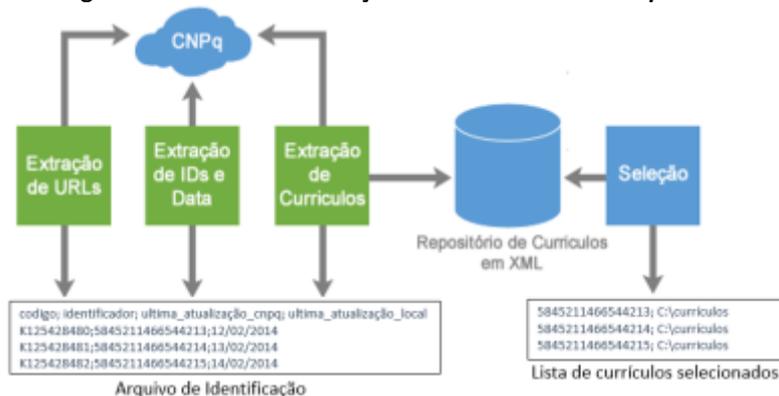
3 DESENVOLVIMENTO

A escolha da Plataforma Lattes como fonte de dados está relacionado: (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 (*eXtensible Markup Language*) foi realizada através da utilização do *LattesDataXplorer*, desenvolvido por Dias (2016) para coletar os dados científicos contido repositório curricular da Plataforma Lattes (Figura 1).

Figura 1: Processo de extração de dados *LattesDataXplorer*.

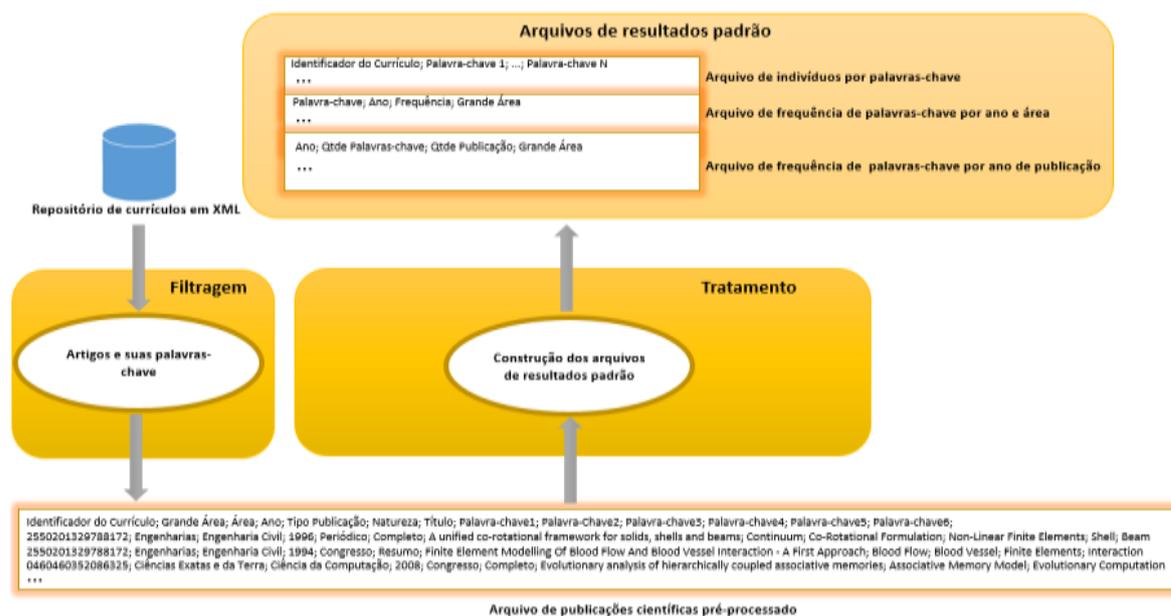


Fonte: Dias (2016)

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).

Em continuação, a Figura 2 apresenta uma visão detalhada do arcabouço de componentes desenvolvido para suportar as análises desejadas. 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: (3.1) limpeza e agrupamento dos dados, (3.2) normalização dos dados e (3.3) elaboração dos arquivos.

3.1 LIMPEZA E AGRUPAMENTO DOS DADOS

Ferraz et. al. (2014) destacam que, analisar palavras-chave de artigos cadastrados na Plataforma Lattes não é uma tarefa trivial, tendo em vista que a escolha das palavras não segue um padrão pré-definido, e, com isso, tem-se uma coleção muito grande de palavras. Na tentativa de contornar este problema, foi desenvolvido um método que realiza o processamento das palavras-chave com o objetivo de excluir termos que não representam tópicos de pesquisas. Além disso, o método agrupa as palavras-chave que foram escritas de formas distintas, mas que possuem mesmo valor semântico.

O método inicia-se obtendo as palavras-chave de cada artigo. Diante disso, cada uma das palavras-chave são associadas ao idioma que foi cadastrado para o artigo, para servir de referência no processo de radicalização. Em continuação, no processo de *lowercase*, todas as palavras são convertidas para minúsculo com a proposta de padronizar o conjunto.

No processo de *stopWords*, são removidos os termos que não possuem valores semânticos. Em seguida, cada uma das palavras-chave passa por um processo de normalização para extrair as letras acentuadas e substituí-las pelo seu equivalente sem acentuação.

O processo de radicalização consiste na redução das palavras-chave a seu radical, e com isso, evita a inclusão de palavras com o mesmo significado de formas distintas. No caso de palavras-chave compostas, este processo é executado em cada termo individualmente, e, em seguida, são concatenados formando uma única palavra. A Tabela 1 apresenta um exemplo de transformação de uma palavra-chave de artigo científico após a execução da etapa de limpeza e agrupamento.

Tabela 1: Exemplo de transformação de uma palavra extraída de um artigo científico

Etapa Algoritmo	Resultado
Recebimento da Palavra-Chave	Gerência de Dados
Deteção de idioma	Gerência de Dados, Português
LowerCase	gerência de dados, Português
StopWords	gerência dados, Português
Normalização	gerencia dados, Português
Radicalização	gerenc dad, Português
	gerenc-dad

Fonte: Elaborado pelos autores

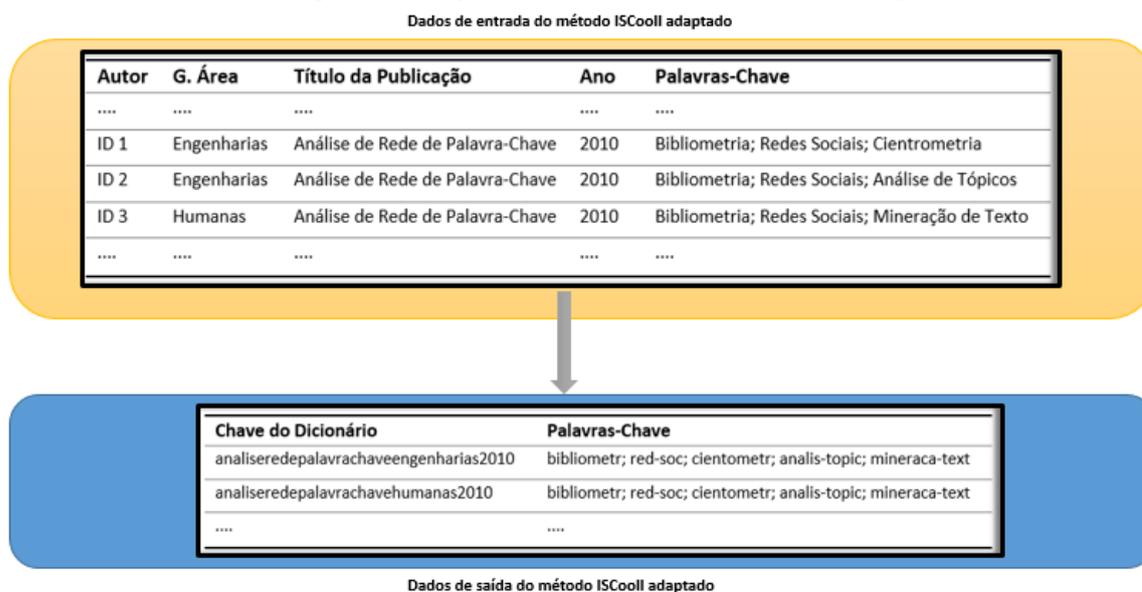
3.2 NORMALIZAÇÃO DOS DADOS

No repositório da Plataforma Lattes, não é uma situação incomum dois pesquisadores cadastrarem um artigo que publicaram juntos utilizando diferentes informações. Em contrapartida, o preenchimento igualitário das informações pode gerar duplicidade de dados quando considerado todo o repositório, impactando em determinadas análises.

Assim, a normalização tem o objetivo eliminar as inconsistências citadas. Para tanto, fez-se necessário uma adaptação no método *ISCool* proposto por Dias e Moita (2015) para identificar colaborações científicas em grande volume de dados. Enquanto o *ISCool* original utiliza um dicionário para vincular os artigos a seus autores, o *ISCool* adaptado adota um dicionário para vincular os artigos ao conjunto união das palavras-chave. Outra diferença é que, enquanto *ISCo* original utiliza o título do artigo concatenado com o ano de publicação,

o *ISCool* adaptado adiciona também a grande área da publicação, ou, em caso de inexistência, a grande área de atuação informada no currículo. A Figura 3 exemplifica o resultado da aplicação do *ISCool* adaptado a um conjunto de artigos científicos em coautoria.

Figura 3: Exemplo de resultado do método *ISCool* adaptado.



Fonte: Elaborado pelos autores

Portanto, ao término do processamento do conjunto de artigos pelo método *ISCool* adaptado, o dicionário contemplará: (1) em casos de publicações em coautoria, onde os autores informaram a mesma grande área, será considerado um único artigo acompanhado com conjunto união das palavras-chave associadas, e (2) em casos de trabalhos em coautoria onde os autores informaram grandes áreas distintas, o número de artigos será de acordo com a quantidade de diferentes grandes áreas informadas pelos seus coautores.

3.3 ELABORAÇÃO DOS ARQUIVOS

A elaboração dos arquivos tem como intuito facilitar as análises e contribuir na diminuição dos dados a serem processados. Para análises mais abrangente, os arquivos de resultados padrão foram divididos em: arquivos que consideram a coautoria (processados pelas etapas de limpeza e normalização) e os que contém todo o conjunto de artigos (processados apenas pela etapa de limpeza). Portanto, 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

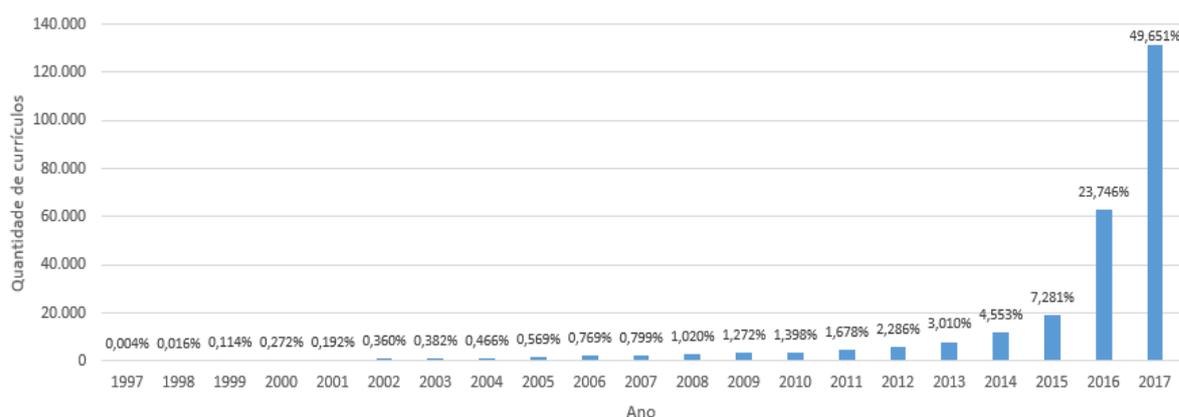
palavras-chave por ano e quantidade de publicação (Ano, frequência palavra-chave, quantidade de publicação e grande área da publicação).

Por fim, vale destacar que a construção dos componentes se deu na linguagem de programação *Python* com a utilização das funcionalidades disponíveis na biblioteca *Natural Language ToolKit (NLTK)*. Neste caso, a *NLTK* contribuiu com os termos referentes as *stopWords* e os algoritmos *Porter Stemmer* para a radicalização de língua inglesa e *Snowball Stemmer* para a língua portuguesa. Para a geração dos resultados apresentados na próxima seção, foram manipulados e processados os arquivos de resultados padrão utilizando *Python* e sua biblioteca *Pandas (para manipulação e análise de dados)*.

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. Para as análises foram considerados os artigos publicados em anais de congressos e em periódicos referentes ao período de 1962 até 2016. A Figura 5 apresenta a distribuição dos currículos dos doutores com base na última data de atualização.

Figura 4: 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 49,651% (131.660) possuem data de última atualização em 2017 e 73,397% (194.626) atualizados nos últimos dois anos. A não atualização de parte desses currículos pode ter motivos variados e de difícil reconhecimento (DIAS, 2016).

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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 necessário 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 apenas um único artigo publicado em coautoria de mesma grande área associado ao conjunto união de suas palavras-chave. Sendo assim, a Tabela 2 apresenta o quantitativo dos dados de artigos extraídos dos currículos dos doutores com data de publicação entre 1962 e 2016.

Tabela 2: Quantitativo de artigos científicos e suas palavras-chave dos currículos dos doutores.

	Tipo de Publicação	Artigos	Palavras-chave vinculadas aos artigos
Geral	Anais de Congresso	8.881.237 (65,4%)	14.284.910 (61,6%)
	Periódicos	4.685.011 (34,6%)	8.889.527 (38,4%)
	Total	13.566.248 (100%)	23.174.437 (100%)
Colaboração	Anais de Congresso	6.776.671 (67,5%)	14.839.228 (61,2%)
	Periódicos	3.263.993 (32,5%)	9.417.084 (38,8%)
	Total	10.040.664 (100%)	24.256.312 (100%)

Fonte: Elaborado pelos autores.

Diante disso, percebe-se que os doutores têm dedicado mais de 60% dos esforços em relação a publicações e preenchimento de palavras-chave em anais de congressos para as duas visões dos dados consideradas. Nota-se também que o número de artigos na visão colaboração é inferior ao respectivo número na visão geral, isso acontece devido a existência de artigos realizados em coautoria cadastrados em diferentes currículos. 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.

Neste ponto, vale ressaltar que o número de doutores representa cerca de 5% da quantidade total de indivíduos da Plataforma Lattes, porém, estes são responsáveis cerca de 70% de todos os artigos cadastrados nos currículos; que, aliado a diversidade e

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

contemporaneidade de atualização destas informações, pondera a validade dos dados para auxiliar na compreensão sobre a evolução da produção científica brasileira (DIAS, 2016).

Entretanto, para as próximas análises foram considerados 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 (24.256.312) das palavras-chave referente a visão colaboração foram processados conforme descrito na seção anterior, e, se reduziu a 2.088.220 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 ao longo do tempo, todas as palavras-chave únicas foram ranqueadas de acordo com a medida de importância de popularidade baseado na frequência a cada período de 5 anos entre 1962 e 2016. Assim, a Tabela 3 apresenta as principais palavras-chave de acordo com sua frequência de utilização em cada quinquênio entre todo o histórico analisado.

Tabela 3: As palavras-chave mais frequentes por quinquênios entre 1962 e 2016.

Ranque	1962-1966	Frequência	1967-1971	Frequência	1972-1976	Frequência
1	Taxonomia	49	Taxonomia	82	Ratos	154
2	Sistemática	30	Doença de Chagas	61	Taxonomia	149
3	Hymenoptera Apoidea	27	Bovinos	58	Bovinos	120
4	Fungos	25	Brasil	54	Cultivo	120
5	Microbiologia	23	Sistemática	53	Soja	109
6	Micologia	23	Cirurgia	49	Brasil	100
7	Brasil	22	Microscopia Eletrônica	43	Doença de Chagas	98
8	Morfologia	22	Coleóptera	37	Glândulas salivares	91
9	Pólen	21	Sedimentologia	37	Milho	91
10	Microscopia Eletrônica	20	Diptera	36	Sementes	91
11	Coleóptera	19	Morfologia	34	Morfologia	90
12	Sol	19	Anatomia	33	Comportamento	88
13	Bovinos	18	Ultraestrutura	31	Adução	87
14	Geologia	18	Abelhas	30	Trypanosoma Cruzi	86
15	Rio Grande do Sul	18	Cardiologia	30	Ecologia	79

Ranque	1977-1981	Frequência	1982-1986	Frequência	1987-1991	Frequência
1	Soja	519	Bovinos	754	Epidemiologia	1.126
2	Bovinos	396	Soja	722	Bovinos	1.056
3	Trypanosoma Cruzi	383	Trypanosoma Cruzi	631	Taxonomia	1.016
4	Sementes	328	Epidemiologia	613	Ratos	1.006
5	Brasil	312	Brasil	581	Brasil	942
6	Ratos	307	Sementes	579	Trypanosoma Cruzi	942
7	Taxonomia	258	Ratos	567	Soja	913

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

8	Milho	255	Taxonomia	506	Educação	895
9	Germinação	238	Milho	505	Amazônia	866
10	Cultivo	219	Doença de Chagas	484	Diagnóstico	854
11	Amazônia	211	Amazônia	480	Sementes	820
12	Ecologia	210	Educação	453	Milho	780
13	Adubação	206	Cultivo	405	Crianças	723
14	Educação	204	Diagnóstico	400	Peixes	716
15	Suínos	203	Peixes	392	Aids	664
Ranque	1992-1996	Frequência	1997-2001	Frequência	2002-2006	Frequência
1	Epidemiologia	3.063	Educação	10.324	Educação	19.680
2	Educação	2.912	Epidemiologia	7.069	Formação do Professor	11.378
3	Bovinos	2.468	Bovinos	5.808	Epidemiologia	10.325
4	Ratos	2.309	Enfermagem	5.799	Crianças	8.290
5	Diagnóstico	2.051	Ratos	5.563	Enfermagem	8.197
6	Brasil	1.923	Formação do Professor	4.950	Bovinos	7.916
7	Amazônia	1.899	Crianças	4.912	Ratos	7.780
8	Aids	1.867	Diagnóstico	4.777	Amazônia	7.232
9	Milho	1.852	Brasil	4.587	Diagnóstico	7.178
10	Sementes	1.814	Milho	4.336	Educação Ambiental	6.911
11	Crianças	1.793	Ensino	4.235	Idoso	6.826
12	Soja	1.710	Amazônia	4.168	Ensino	6.688
13	Trypanosoma Cruzi	1.645	Sementes	3.856	Avaliação	6.140
14	Taxonomia	1.562	Avaliação	3.663	Políticas Públicas	6.133
15	Enfermagem	1.534	Soja	3.594	Brasil	6.001
Ranque	2007-2011	Frequência	2012-2016	Frequência		
1	Educação	23.078	Educação	18.205		
2	Formação do Professor	15.850	Formação do Professor	12.427		
3	Enfermagem	12.910	Enfermagem	11.078		
4	Epidemiologia	10.609	Políticas Públicas	8.473		
5	Políticas Públicas	9.885	Sustentabilidade	7.151		
6	Idoso	9.212	Epidemiologia	7.057		
7	Crianças	7.722	Idoso	6.681		
8	Educação Ambiental	7.344	Ensino	5.832		
9	Biodiesel	7.250	Amazônia	5.744		
10	Amazônia	6.924	Gênero	5.341		
11	Cultura	6.442	Educação Ambiental	5.185		
12	Gênero	6.434	Crianças	5.133		
13	Ensino	6.405	Brasil	4.819		
14	Sustentabilidade	6.345	Memória	4.775		
15	Meio Ambiente	6.287	Biodiesel	4.698		

Fonte: Elaborado pelos autores

Diante disto, verifica-se a evolução do quantitativo das principais palavras-chave a cada quinquênio, isso devido ao aumento de esforços nos respectivos temas, que, possivelmente seguiu o mesmo comportamento do crescimento da produção científica brasileira. Também percebe-se que grande parte dos principais tópicos discutidos no quinquênio corrente, estiveram presentes no quinquênio anterior. Isso mostra que os interesses centrais dos doutores vão mudando gradualmente ao longo do tempo, que, pode estar relacionado a uma demanda da sociedade ou a maturação de um determinado tema.

Por exemplo, “Aids” apareceu entre as principais palavras-chave nos quinquênios de 1987-1991 e 1992-1996, possivelmente reflexo do impacto que a doença causou no início da década de 1980. Coincidentemente, nestes dois quinquênios foi registrado o maior percentual de similaridade entre os principais tópicos estudados, onde apenas se diferiram entre “Peixes” e “Enfermagem”. Nesse caso, “Peixes” foi frequentemente utilizado em 1987-1991, dando

**XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP**

lugar a “Enfermagem” em 1992-1996, que se manteve em foco em todos os quinquênios posteriores. Uma das hipóteses para tal acontecimento, é o processo de regulamentação do exercício profissional que reconheceu as categorias de enfermeiro durante a década de 1980.

Contudo, notadamente a palavra-chave “Educação” possui destaque absoluto na pesquisa nacional. Entre as principais, destacou-se pela primeira vez em 1977-1981. Em 1997-2001 alcançou o topo em popularidade, de onde não saiu até 2012-2016. As palavras “Formação do Professor” e “Ensino”, ranqueadas entre as primeiras no período de 1997-2001 até 2012-2016, corroboram à dedicação em pesquisas ligadas à educação. Ainda, nota-se que várias outras palavras-chave relacionadas a saúde, agropecuária, meio ambiente e política são frequentemente utilizadas ao longo do tempo. Por fim, destaca-se a palavra-chave “Brasil”, que exceto em 2007-2011, figurou-se em popularidade em todos os outros quinquênios.

Ao contrário da medida de frequência, Ohniwa et. al. (2010) propuseram um método que mede a taxa de aparecimento dos termos, e, assim, definem os primeiros 5% como emergentes. A taxa de cada termo é calculada da seguinte forma: considerando o ano A de um respectivo termo, a taxa de aparecimento deste termo é a soma de sua frequência nos anos $A + 1$ e $A + 2$ dividido pela soma de sua frequência nos anos $A - 1, A, A + 1$ e $A + 2$. Com isso, este método foi aplicado ao conjunto de palavras-chave únicas para destacar as principais taxas de aparecimento em cada quinquênio e, assim, permitir a comparação com as palavras-chave ranqueadas pela frequência. A Tabela 4 apresenta as palavras-chave que mais emergiram em cada quinquênio entre 1962 e 2016.

Tabela 4: As palavras-chave mais cresceram de utilização entre 1962 e 2016.

Ranque	1962-1966	Emergente	1967-1971	Emergente	1972-1976	Emergente
1	Schistosoma mansoni	3,50	América do Sul	3,60	Panstrongylus Megistus	3,78
2	Produtos	3,50	Taxonomia Vegetal	3,60	Músculo Masseter	3,75
3	Bivalves	3,43	Distúrbios Hidroeletrólitos	3,57	Geologia Costeira	3,73
4	Comportamento	3,43	Água	3,56	Cirurgia Pediátrica	3,69
5	Adubação Mineral	3,40	Alho	3,56	Anticorpos	3,67
6	Sol	3,40	Espectroscopia Mössbauer	3,50	Crítica	3,67
7	Arenito	3,33	Psiquiatria	3,50	Sêmen	3,63
8	Parasita	3,33	Distrito Federal	3,50	Centro de Cores teoria	3,63
9	Madeira	3,33	Xisto	3,50	Acetonitrila	3,62
10	Tetraciclina	3,33	Controle de Vetores	3,50	Análise de Sistemas	3,60
11	Aprendizagem	3,33	Polpa Dental	3,50	Hemoglobinopatia	3,60
12	Genética	3,33	Hormônio	3,50	Dióxido de enxofre (so2)	3,60
13	Uso do Solo	3,33	Filosofia da matemática	3,50	Deficiência Mineral	3,60
14	Cardiopatia Chagásica	3,33	Estenose Pulmonar	3,50	Fitovírus	3,60
15	Enzimas proteolíticas	3,33	Ditassa	3,50	Linguística	3,57
Ranque	1977-1981	Emergente	1982-1986	Emergente	1987-1991	Emergente
1	Barbatimão	3,87	Peixes Brasileiros	3,85	Desenvolvimento Máquinas	3,85
2	Salsão	3,71	Cultivo de Microalgas	3,82	Proteína texturizada	3,82
3	Tropical Soils	3,67	Urease	3,80	Hemilophini	3,80
4	Terminação	3,64	Compression	3,80	Saúde da Mulher	3,78
5	Varrotose	3,64	Tântalo	3,75	Trichoderma Harzianum	3,77

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

6	Inhibition	3,63	Sphincter	3,75	Santana do Riacho	3,77
7	Objetivos de Ensino	3,60	Bromopride	3,75	Anemia de Fanconi	3,75
8	Aberrações Cromossômicas	3,60	Pressure	3,71	Trombina	3,75
9	Ontogênese	3,60	Endocytosis	3,71	Raleio de Frutos	3,75
10	Cortisol	3,60	Poconé	3,70	Methylmalonate	3,73
11	Máquinas Agrícolas	3,60	Imunidade Passiva	3,69	Ajuste Oclusal	3,71
12	Época de Corte	3,60	Amido de Mandioca	3,67	Leishmaniose Visceral	3,71
13	Saúde Ocupacional	3,60	Territorialidade	3,67	Veículo Automotivo	3,71
14	Doença de Newcastle	3,60	Bioimpedância	3,67	Excreção de Eletrólitos	3,71
15	Flavonoides	2.938	Educação Especial	8.620	Administração	4.279
Ranque	1992-1996	Emergente	1997-2001	Emergente	2002-2006	Emergente
1	Folículos Pré-Antrais	3,92	Antioxidant Capacity	3,88	Agronegócio cooperativo	3,87
2	Anastrepha Fraterculus	3,90	Ovos de Codorna	3,88	Tenoplastia	3,87
3	Cabro Pré-Reunido	3,86	Microarrays	3,87	Ceraeochrysa Cubana	3,86
4	Material Dentário	3,85	Vaca Seca	3,87	Propriedades Estado Sólido	3,86
5	Toughening	3,83	Gestão em Microempresa	3,87	Estigma e Discriminação	3,86
6	Reflexão Computacional	3,82	Vidros de Óxidos	3,86	Transferência de genes	3,86
7	Área da Saúde	3,82	C-pseudotuberculosis	3,85	Produção de Material	3,86
8	Astrolabe	3,82	Bombeiros	3,85	Design for Assembly	3,85
9	Engenharia de Sistemas	3,82	Universidade Corporativa	3,83	Engenharia Offshore	3,83
10	Ferrugem Tropical	3,82	Altura do Pasto	3,83	kava	3,83
11	Psicoses Infantis	3,80	Coefficientes de Similaridade	3,83	Músculo Papilar Isolado	3,82
12	Morfopedologia	3,78	Nutrient Transport	3,83	Estabilizadores da Escápula	3,82
13	Clorose Variegada dos Citros	3,78	Infecção Assintomática	3,83	Poncirus Trifoliata	3,82
14	Tape casting	3,78	Low Latitude Ionosphere	3,82	Hipervalentes de Telúrio	3,82
15	Coi	3,78	Supported Metalocene	3,82	Jatropha Curcas	3,81
Ranque	2007-2011	Emergente	2012-2016	Emergente		
1	Urochloa brizantha	3,91	Operador de Consequência	3,91		
2	Sociologia Rural	3,89	Camada Porosa de Atrito	3,83		
3	Avaliação Sistema de Saúde	3,86	Microcephaly	3,83		
4	Olfactory Ensheathing Cells	3,85	Paper Based Microfluidics	3,81		
5	Diagnóstico Automático	3,85	Maturity Stages	3,80		
6	Artery Mesenteric	3,85	Sistema Alagado Construído	3,80		
7	Plathymenia	3,85	Delimitação Automática	3,80		
8	Comuna de Paris	3,84	Adsorbents	3,79		
9	Salas de Multi-Recursos	3,83	Provisionamento	3,78		
10	Especiação de Metais	3,83	Cirurgia Nasal	3,78		
11	Jejum Hídrico ou Alimentar	3,83	Bovinos	3,78		
12	Modelo de Resposta Gradual	3,82	Cytogenetics	3,78		
13	Radiologia Ambiental	3,82	Indissociabilidade	3,77		
14	Internet do Futuro	3,81	Crustal Structure	3,75		
15	Produto Educacional	3,80	Camarão Branco do Pacífico	3,75		

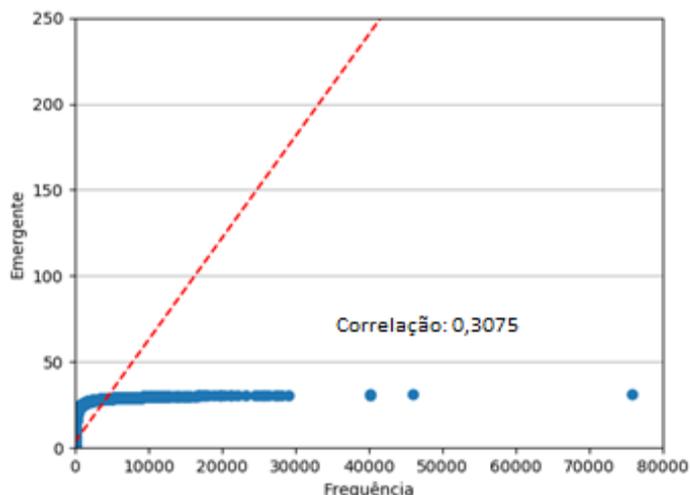
Fonte: Elaborado pelos autores

É possível verificar que, diferentemente do ranqueamento realizado pela frequência, as palavras-chave que apresentaram maior taxa de aparecimento no quinquênio corrente não possui nenhuma relação com as palavras-chave que estiveram entre as principais no quinquênio anterior. Ou seja, a cada 5 anos os principais interesses dos doutores brasileiros se renovavam por completo, não possuindo nenhuma relação passada. Também observa-se que, ao comparar o conjunto das palavras-chave mais frequentes com as palavras que tiveram maior taxa de aparecimento por quinquênio ao longo dos anos, foi verificado que apenas a palavra-chave “Sol” que ocorreu no período de “1962-1966” apareceu entre as principais de ambas medidas de importância.

Diante disto, para melhor compreender a relação entre os resultados das medidas de importâncias utilizadas, uma análise de correlação foi realizada. Para tal análise, como

apresentado na Figura 5, foi calculado o coeficiente dos postos de Spearman considerando todo o conjunto histórico das 2.088.220 palavras-chave únicas ranqueadas pela medida de importância baseada em frequência e taxa de aparecimento, visando verificar a existência de uma relação direta entre os resultados apresentados pelas medidas de importância utilizada.

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



Fonte: Elaborado pelos autores

Por fim, como pode ser observado, o resultado apresenta um coeficiente de correlação positivo de 0,3075 entre frequência e taxa de aparecimento, o que significa que, apesar de positiva, a correlação é fraca entre as medidas de importâncias de tópicos utilizadas.

5 CONSIDERAÇÕES FINAIS

Este estudo contribuiu para a identificação dos tópicos de pesquisas que obtiveram maior interesse em cada época por parte dos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes. 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. Adicionalmente, ao comparar as principais palavras-chave pela medida de frequência com a medida de taxa de aparecimento por período, foi possível constatar uma baixa similaridade, e, uma baixa correlação considerando todo o conjunto de análise.

Como contribuição à validade dos resultados, é possível destacar a relevância do repositório da Plataforma Lattes para estudos bibliométricos sobre o desenvolvimento da

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

ciência brasileira, aliado a abrangência e contemporaneidade de atualização dos dados considerados para análises.

Entretanto, a análise das palavras-chave utilizando apenas medida de importância de popularidade baseado em frequência sofre influência quantitativa, e, no caso da Plataforma Lattes que possui artigos publicados em conferências e periódicos com diferentes níveis de qualidades, análises apenas quantitativas podem induzir a desconsideração de fatores qualitativos. Com isso, como trabalhos futuros, espera-se que sejam realizadas análises que consideram características qualitativas apropriadas ao contexto da Plataforma Lattes, fatores temporais para determinar a relevância de um tópico em cada época, e, análises baseadas em técnicas para análises de redes sociais para determinar as palavras-chave mais importante da rede. 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.

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.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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.

XVIII ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO – ENANCIB 2017
23 a 27 de outubro de 2017 – Marília – SP

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.