

Visualização analítica das palavras-chaves nos eventos científicos: proposta a partir do Currículo Lattes

Jether Oliveira Gomes

Mestre e Doutorando em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), <http://lattes.cnpq.br/6063087854331543>, jetherog@gmail.com

Thiago Magela Rodrigues Dias

Docente e pesquisador do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Doutor em Modelagem Matemática e Computacional pelo Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), <http://lattes.cnpq.br/4687858846001290>, thiagomagela@gmail.com

Adilson Luiz Pinto

Docente e pesquisador da Universidade Federal de Santa Catarina (UFSC), Doutor em Documentación pela Universidad Carlos III de Madrid (UC3M), Espanha, <http://lattes.cnpq.br/4767432940301118>, adilson.pinto@ufsc.br

Gray Farias Moita

Docente e pesquisador do Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG), Doutor em Aeronáutica pelo Imperial College of Science, Technology and Medicine (ICSTM), Inglaterra, <http://lattes.cnpq.br/2550201329788172>, gray@dppg.cefetmg.br

Submetido em: 22/08/2017. Aprovado em: 24/10/2017. Publicado em: 22/02/2018.

RESUMO

A produção de trabalhos científicos apresentou crescimento importante nas últimas décadas, sendo a internet seu principal fator de acesso e difusão. Consequentemente, nota-se um esforço global de todas as áreas do conhecimento quanto a estudos sobre dados de publicações científicas, a fim de conhecer o que tem sido pesquisado. Tais estudos podem servir a diversos propósitos, como fornecer base para construção de políticas de incentivo à pesquisa visando novos avanços na ciência. O objetivo deste trabalho é identificar e analisar os principais tópicos de pesquisas publicados em cada grande área do conhecimento, durante a trajetória da ciência brasileira por pesquisadores doutores que possuem currículos cadastrados na Plataforma Lattes. Para isso, após a aquisição dos currículos dos doutores, foi desenvolvido um arcabouço de componentes responsável por filtrar, tratar e consolidar separadamente os dados das publicações científicas a serem analisados. Dentre esses dados, estão todas as palavras-chave de artigos publicados entre 1962 e 2016 em anais de congressos e em periódicos, que são ranqueadas em cada grande área do conhecimento através da aplicação da medida de popularidade de tópico baseado em frequência. Os resultados possibilitaram uma caracterização geral das palavras-chave utilizadas pelos doutores, e assim, identificar os principais tópicos de pesquisas desenvolvidos em cada grande área do conhecimento. Como considerações finais, observou-se a sinergia entre as grandes áreas do conhecimento quanto aos principais tópicos de investigação e também a contribuição de cada grande área em quantitativo de artigos publicados.

Palavras-chave: Palavras-chave. Tópicos de pesquisas. Bibliometria. Plataforma Lattes.

Analytical visualization of the keywords in scientific meeting: proposed from the Lattes platform

ABSTRACT

The production of scientific papers has presented important growth in the last decades, the Internet being their main factor of access and diffusion. Consequently, there is a global effort of all fields of knowledge regarding studies of scientific publications data, in order to know what has been researched. Such studies can serve a number of purposes, such as providing a basis for developing policies to stimulate research aiming at new advances in science. In this context, the objective of this work is to identify and analyze the main topics of research published in each major field of knowledge during the trajectory of Brazilian science by PhD researchers who have curricula registered in the Lattes Platform. To this end, after the acquisition of the doctors' curricula, a framework of components was developed responsible for filtering, treating and consolidating separately the data of the scientific publications to be analyzed. Among these data are all article keywords published between 1962 and 2016 in congress annals and journals that are ranked in every major field of knowledge through the application of frequency-based topic popularity measure. The results allowed a general characterization of the keywords used by the graduates, and thus, to identify the main topics of research developed by each major area of knowledge. As final considerations, we observed the synergy between the major areas of knowledge regarding the main topics of research and also the contribution of each large area in quantitative of published articles.

Keywords: Keywords. Research topics. Bibliometrics. Lattes Platform.

Visualización analítica de las palabras-clave en los eventos científicos: propuesta a partir del currículo Lattes

RESUMEN

La producción de trabajos científicos ha presentado un crecimiento importante en las últimas décadas, siendo Internet el principal factor de acceso y difusión de estos. En consecuencia, se nota un esfuerzo global de todas las áreas del conocimiento en cuanto a estudios sobre datos de publicaciones científicas, a fin de conocer lo que ha sido investigado. Tales estudios pueden servir a diversos propósitos, como proporcionar base para la construcción de políticas de incentivo a la investigación buscando nuevos avances en la ciencia. En este contexto, el objetivo de este trabajo es identificar y analizar los principales tópicos de investigaciones publicados en cada gran área del conocimiento durante la trayectoria de la ciencia brasileña por investigadores doctores que poseen currículos registrados en la Plataforma Lattes. Para ello, después de la adquisición de los currículos de los doctores, se desarrolló un marco de componentes responsable por filtrar, tratar y consolidar por separado los datos de las publicaciones científicas a ser analizadas. Entre estos datos se incluyen todas las palabras clave de artículos publicados entre 1962 y 2016 en anales de congresos y en periódicos que se clasifican en cada gran área del conocimiento a través de la aplicación de la medida de popularidad de tópico basada en frecuencia. Los resultados posibilitar una caracterización general de las palabras clave utilizadas por los doctores, y así identificar los principales tópicos de investigación desarrollados por cada gran área del conocimiento. Como consideraciones finales, se observó la sinergia entre las grandes áreas del conocimiento en cuanto a los principales tópicos de investigación y también la contribución de cada gran área en cuantitativo de artículos publicados.

Palabras clave: Palabras clave. Temas de investigación. Bibliometría. Plataforma Lattes.

INTRODUÇÃO

O grande número de informações disponibilizadas pela internet e a sociabilização da atividade científica por parte de redes de pesquisadores são os fatores essenciais para o atual desenvolvimento da ciência (BRITO et al., 2016). Serviços como bibliotecas digitais e sítios para registro individual de produção científica são alguns exemplos de como a internet tem contribuído quanto ao número de trabalhos publicados. Os trabalhos publicados podem ser acessados instantaneamente contribuindo para a expansão do conhecimento (DIAS, 2016).

Para Carneiro (2003), o conhecimento é o principal elemento para a geração do desenvolvimento. A divulgação científica contribui para popularização do conhecimento e aproxima o cidadão comum dos benefícios que tem direito de reivindicar para a melhoria do seu bem-estar social, dando-lhe uma visão mais clara sobre os problemas que enfrenta no dia a dia.

Numa sociedade competitiva, implementar o conhecimento técnico e científico é tarefa primordial para alavancar o desenvolvimento econômico e social (CAVACINI, 2016). Porém, muitas das vezes, seu plano de implementação é condicionado à existência de recursos limitados e a imposição cada vez maior de racionalidade e objetividade na aplicação dos poucos recursos disponíveis.

Tão importante quanto ter investimentos é ter habilidade para controlar, entender e medir o nível científico das nações e/ou grupos individualizados, negócios e fundações que devem decidir suas prioridades científicas. Portanto, torna-se imprescindível estudar a produção científica para a implementação desse conhecimento, superação das dificuldades e alcance dos pressupostos de racionalidade e objetividade (SAES, 2005).

Diante disso, pesquisadores de todos os domínios têm dedicado esforços com intuito de analisar a produção científica sobre diferentes perspectivas, como, por exemplo, a partir da análise de tópicos extraídos de publicações científicas.

Neste caso, um tópico pode ser entendido como termo que representa um dos assuntos associados a determinado documento (BORGES et al., 2015).

Assim, trabalhos que analisam tópicos de pesquisa constituem formas de melhorar a compreensão do comportamento de grupos de pesquisadores, bem como evidenciar o que tem sido produzido acerca da ciência. Adicionalmente, outras informações podem ser descobertas, a saber: (1) principais tópicos em discussão, (2) tópicos novos, e (3) tendências de pesquisas. A partir destas descobertas, ainda é possível realizar análises baseadas em agrupamento, ranqueamento e recomendação. Trabalhos dessa natureza também funcionam como uma boa revisão da literatura, o que permite, por exemplo, a verificação por parte do setor industrial se o que está sendo desenvolvido pela ciência contempla as necessidades da indústria (KHAN; WOOD, 2015).

Para tanto, tais trabalhos geralmente utilizam títulos e resumos de publicações científicas para extrair tópicos de pesquisas e analisá-los através de aplicação de métodos bibliométricos ou técnicas baseadas em redes sociais ou análises de tendências. Entretanto, títulos e resumos podem não representar de maneira efetiva os principais assuntos abordados, devido à necessidade do autor se preocupar com a estrutura semântica dos termos em suas descrições. Logo, uma abordagem interessante é a análise das palavras-chave de publicações científicas, pois são inseridas cuidadosamente 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; YI; CHOI, 2012; KHAN; WOOD, 2015). Neste trabalho, palavras-chave de artigos científicos são igualmente referenciadas como tópicos de pesquisas.

Dentre os trabalhos encontrados na literatura, geralmente as análises realizadas utilizam um conjunto restrito de dados advindos de repositórios internacionais, que são específicas de determinada área ou periódico. No entanto, por tratarem de análises específicas e utilizarem de repositórios internacionais, não podem

representar de modo 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 tarefa relevante para a compreensão da ciência brasileira.

Nesse contexto, para análises sobre o desenvolvimento científico brasileiro, o repositório de dados da Plataforma Lattes é tido como um diferencial (LANE, 2010). O repositório é composto por dados de grupos de pesquisas, instituições e currículos de mais de 5 milhões de indivíduos (DIAS, 2016). Também é fonte de informação para órgãos que avaliam o Sistema Nacional de Pós-Graduação do Brasil, e agências de fomento que financiam pesquisas e ofertam bolsas de estudos. Contudo, mesmo estando disponível livremente na internet, esses dados ainda não foram amplamente analisados (DIGIAMPIETRI, 2015). Dos trabalhos que exploram os 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, este trabalho tem como objetivo realizar uma análise sobre as palavras-chave de artigos publicados entre 1962 e 2016 em anais de congressos e em periódicos pelos doutores que têm currículos cadastrados na Plataforma Lattes. O estudo contempla a extração e processamento das palavras-chave, para, posteriormente, serem realizadas análises bibliométricas, a fim de identificar e mapear os tópicos de interesses de cada grande área do conhecimento, evidenciando-os que possuem destaque para medir o relacionamento entre elas quanto aos principais tópicos de investigados.

TRABALHOS CORRELATOS

Esforços para identificar e analisar tópicos de pesquisa constitui um jeito 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 no contexto científico.

De acordo com Choi et al. (2011) e Khan e Wood (2015), os estudos podem ser baseados tanto em termos extraídos dos títulos das publicações quanto das palavras-chave de produções bibliográficas.

Por exemplo, Vinkers et al. (2015) estudaram os resumos e títulos de artigos científicos publicados entre 1974 e 2014 extraídos da PubMed, com o intuito de investigar as tendências de utilização de termos positivos, negativos e neutros. Para tanto, definiram conjuntos de termos para cada classe, para, em seguida, serem quantificados por período de tempo e comparados com os livros cadastrados no Google Books Ngram Viewer. Os resultados mostraram que os resumos e os títulos estão cada vez mais contendo termos positivos, o que provavelmente está relacionado à evolução da literatura científica. Contudo, os autores ressaltaram que tal fato pode estar ligado ao exagero dos pesquisadores na utilização de uma linguagem positiva perante seus resultados.

Pollack e Adler (2015) realizaram análises bibliométricas para descobrir tendências de pesquisas na área de gerenciamento de projetos a partir de 94.472 artigos publicados entre 1962 e 2012, cadastrados nos repositórios da Scopus e Web of Science. 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 coocorrência para identificar os tópicos emergentes.

Em Zhu et al. (2015), os autores analisaram as palavras-chave de 363.458 teses, publicadas entre 1986 e 2014, referentes à indústria de petróleo e gás cadastradas no repositório do repositório da China National Knowledge internet (CNKI), para entender o processo de desenvolvimento da área e evidenciar os assuntos atuais. Inicialmente, realizaram a mineração dos dados e, em seguida, analisaram as frequências, coocorrências e as redes das palavras-chave.

Já no trabalho de Chen et al. (2015), os autores analisaram palavras-chave de artigos com o propósito de evidenciar o foco de pesquisa das universidades da China na área de ciência da informação.

Para tal, propuseram e utilizaram o método KAI (Keyword Activity Index) para identificar os principais tópicos das 8 melhores universidades, que foi representada por um total de 277.721 palavras-chave referentes a 65.653 artigos publicados no período de 2000 a 2013 registrados no repositório Full Text.

Em Silva et al. (2016), os autores propuseram uma metodologia baseada em técnicas de redes sociais e análises de textos para selecionar artigos mais relevantes de determinada área do conhecimento e mapear sua organização. Primeiramente, os artigos mais relevantes foram destacados através da utilização de redes de citações, onde cada artigo é tratado como um nó da rede e cada citação de um artigo por outro é considerada uma conexão. Em seguida, foram realizadas análises dos termos extraídos dos títulos e resumos das publicações com intuito de destacar os termos mais importantes para rotular as comunidades existentes.

Hong et al. (2016) realizaram a combinação de análises bibliométricas, coocorrência de termos e técnicas de análises de redes sociais para identificar a estrutura do conhecimento e a evolução dos tópicos de pesquisas relacionados a médicos que atuaram como clínicos gerais no período de 1999 a 2014. Os dados para estudos foram coletados no repositório da PubMed, sendo 10.704 publicações e seus 42.695 termos *Mesh* associados. Estes dados foram agrupados e analisados por faixas de tempo, para facilitar a verificação da evolução dos tópicos de pesquisas.

Vários outros trabalhos têm analisado publicações científicas para compreensão sobre a evolução da ciência com diferentes propósitos, como taxas de carbono (ZHANG et al., 2016), gestão na América Latina (RONDA-PUPO, 2016); linhas de produtos de software (HERADIO et al. 2016), tendências de pesquisas (RAVIKUMAR et al., 2015); epidemiologia na Alemanha (PETER et al., 2016); saúde no Brasil (PEREIRA et al., 2007); repercussão do acidente de Chernobyl (MRYGLOD et al., 2016); entre outros.

Entretanto, é importante destacar que os trabalhos referenciados até o momento utilizam dados de repositório internacionais (determinada área ou periódico). A seguir, são apresentados os poucos e recentes trabalhos encontrados que analisam tópicos de pesquisas e que utilizam fontes de dados nacionais (mais especificamente dados da Plataforma Lattes).

Em Medeiros e Mena-Chalco (2013), os autores analisaram 633.508 currículos e 4.946.990 publicações para estudar a rede social dos indivíduos de áreas distintas. Adicionalmente, calcularam as frequências dos termos dos títulos para identificar quais são mais utilizados em cada época. Em continuação, utilizaram mapas de termos e árvores de mapas para as 200 palavras mais frequentes de cada área por período de tempo.

O trabalho de Digiampietri et al. (2014) apresentou uma análise sobre a rede social formada a partir 2002 currículos de indivíduos que atuam no Brasil em inteligência artificial ou inteligência computacional. Além disso, analisaram a frequência relativa dos termos contidos nos títulos dos artigos publicados entre 1993 a 2012 para ilustrar a evolução dos temas abordados ao longo de cada ano.

Trucolo e Digiampietri (2014a) aplicaram regressões nas medidas 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 de pesquisas. Assim, um termo foi considerado tendência se possuísse alta previsão de *TF-IDF*. Em Trucolo e Digiampietri (2014b), os autores aplicaram as mesmas regressões nos termos de 57.501 títulos de artigos publicados entre 1911 e 2011 pelos professores de ciência da computação. Em continuação, Trucolo (2016) considerou a fonte geradora da informação, e com isso, utilizou também resultados de métricas de redes sociais para análises de tendências nos mesmos dados estudados em Trucolo e Digiampietri (2014b).

Dentre os trabalhos apresentados, verificou-se que a estratégia de analisar termos extraídos dos títulos das publicações científicas tem recebido grande atenção dos estudos (no caso da Plataforma Lattes, são todos). Embora esta estratégia tenha validade, claramente possui suas limitações, uma vez que nem sempre os títulos conseguem expressar de modo efetivo todo o conteúdo abordado pelo trabalho. Com isso, a análise das palavras-chave que são associadas aos artigos vem ganhando força na literatura, pois foram escolhidas cuidadosamente pelos próprios autores com o foco principal de evidenciar os tópicos centrais que permeiam o trabalho de maneira objetiva.

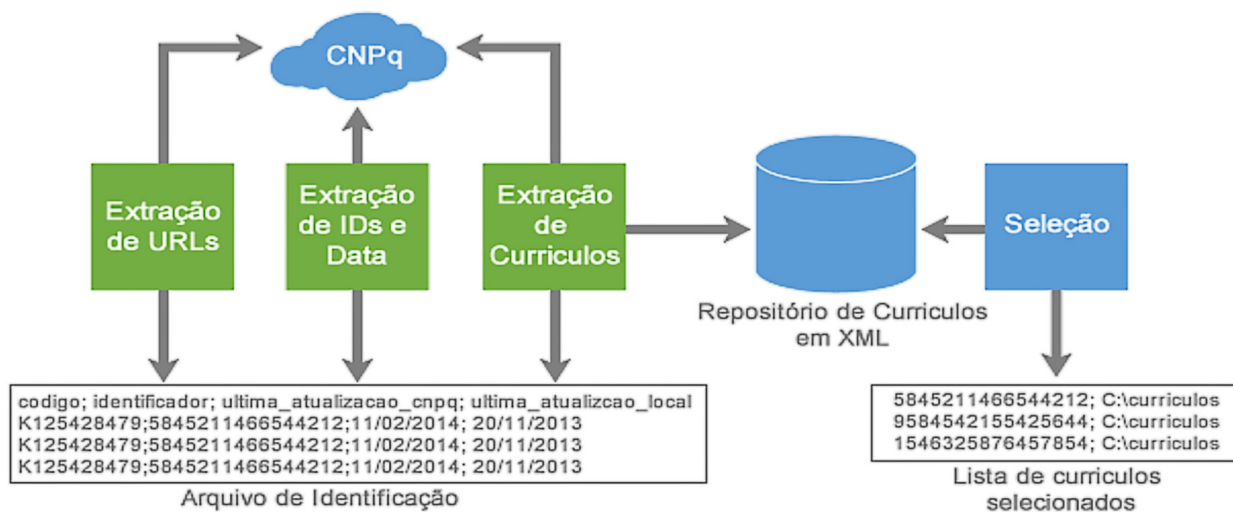
Diante do exposto, o presente trabalho apresenta-se como a análise abrangente sobre as palavras-chave dos artigos publicados pelos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes para destacar os principais tópicos de pesquisas de cada grande área do conhecimento e verificar a relação entre elas.

MATERIAIS E MÉTODOS

A motivação escolha da Plataforma Lattes como fonte de informação está relacionada basicamente a quatro fatores: (i) aos dados estarem disponíveis na internet e não terem sido amplamente analisados (DIGIAMPIETRI, 2015); (ii) tratar da integração de dados de produções científicas de todas as áreas de C&T existentes na ciência brasileira ao longo de toda trajetória; (iii) 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, (iv) por ser uma poderosa fonte para fornecimento de dados de alta qualidade para medir e avaliar o desempenho acadêmico nacional (LANE, 2010).

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 nos currículos cadastrados na Plataforma Lattes (figura 1).

Figura 1 – Processo de aquisição dos dados LattesDataXplorer



Fonte: Dias (2016)

Inicialmente, todos os currículos cadastrados na Plataforma Lattes são extraídos através da execução de três etapas pelo LattesDataXplorer, a saber: (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. Para, posteriormente, através do módulo de seleção, selecionar dentre estes, os currículos que possuem nível de capacitação doutorado concluído, definindo assim o conjunto principal de arquivos XML a serem processados por este trabalho.

Em continuação, apesar de os currículos XML da Plataforma Lattes permitirem o cadastramento de várias outras informações, neste trabalho optou-se por utilizar apenas informações de artigos científicos publicados em anais de congressos e em periódicos, com o intuito de assim representar a produção científica brasileira. A justificativa se dá pelo fato dos artigos serem considerados o principal caminho de disseminação de novos conhecimentos na maioria das disciplinas científicas (RONDA-PUPO, 2015).

A figura 2 apresenta uma visão geral do arcabouço de componentes desenvolvidos que suporta as análises desejadas.

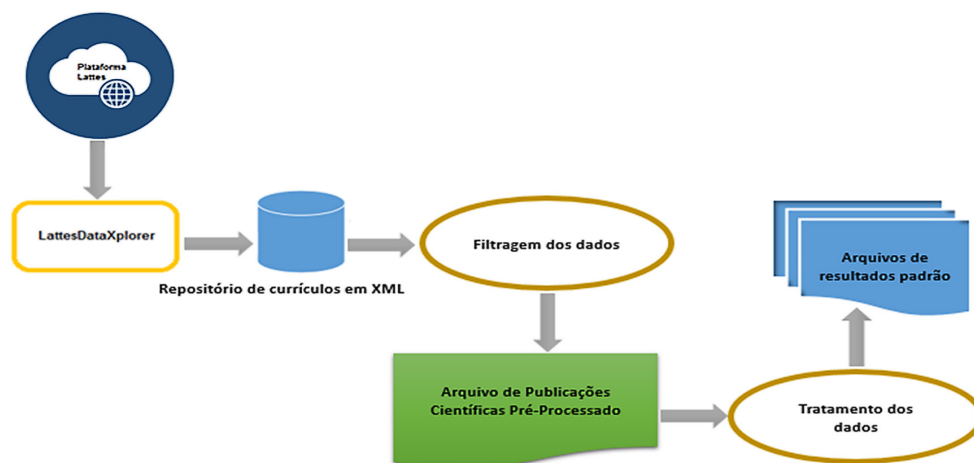
Nela, os componentes “filtragem dos dados” e “tratamento dos dados” são os responsáveis pelo processo de seleção e tratamento dos dados dos currículos a serem analisados, e, concomitantemente, diminuir o tempo de processamento computacional.

Os componentes de “filtragem dos dados” e “tratamento dos dados” são ilustrados com maiores detalhes na figura 3.

De posse dos currículos, o componente de “filtragem dos dados” realiza a etapa de mineração nos arquivos XML para extrair as informações dos artigos publicados em anais de congressos e em periódicos, 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 a realização das análises. Esse componente realiza basicamente três etapas, a saber: limpeza e agrupamento dos dados, normalização dos dados e elaboração dos arquivos.

Figura 2 – Visão geral do conjunto de componentes



Fonte: Elaborado pelos autores

Figura 3 – Processo de filtragem e tratamento dos dados



Fonte: Elaborado pelos autores.

LIMPEZA E AGRUPAMENTO DOS DADOS

Tendo em vista que o cadastramento das palavras-chave de um artigo científico no repositório curricular da Plataforma Lattes é de responsabilidade dos respectivos pesquisadores, e isso é feito livremente por eles, o que significa que pode ser inserido qualquer conjunto de caracteres como uma palavra-chave. A partir disto, geralmente tem-se uma coleção muito grande de palavras-chave sem nenhum padrão predefinido. Na tentativa de contornar esse 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* (*stemming*). Em continuação, no processo de *lowercase*, todas as palavras são convertidas para minúsculo. 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 os acentos e caracteres especiais. E, por último, o processo de radicalização consiste na redução das palavras-chave a seu radical, para evitar a inclusão de palavras com o mesmo significado de formas distintas. No caso de palavras-chave compostas, o processo é executado em cada termo individualmente e, em seguida, são concatenados formando uma só palavra.

A tabela 1 apresenta exemplo de transformação de palavra-chave de artigo científico após a execução da etapa de limpeza e agrupamento.

Tabela 1 – Exemplo de transformação de palavra extraída de artigo científico

Etapa Algoritmo	Resultado
Recebimento da Palavra-Chave	Gerência de Dados
Detecção 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

NORMALIZAÇÃO DOS DADOS

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

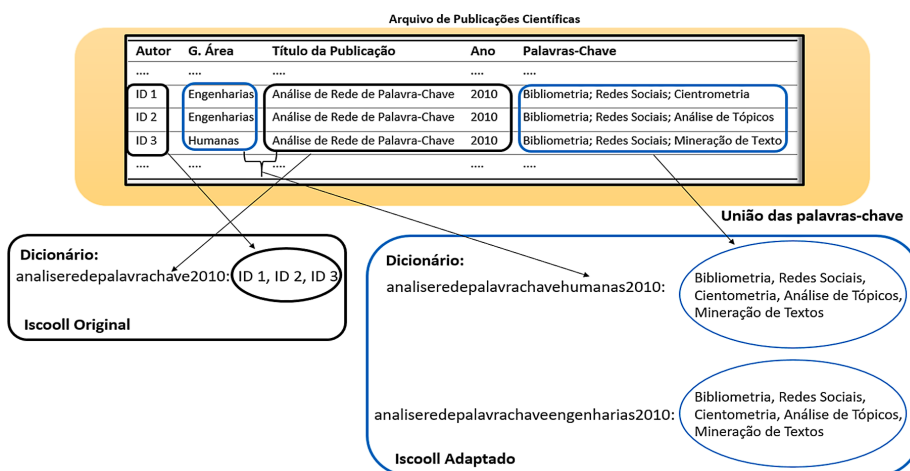
Assim, a normalização tem o objetivo eliminar as inconsistências citadas. Para tanto, fez-se necessário uma adaptação no método ISCooll proposto por Dias e Moita (2015) para identificar colaborações científicas em grande volume de dados. Enquanto o ISCooll original utiliza um dicionário para vincular 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 figura 4 exemplifica o resultado da aplicação do ISCooll original e adaptado a artigo científico realizado em coautoria.

Portanto, ao término do processamento do conjunto de artigos pelo método ISCooll 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 em que os autores informaram grandes áreas distintas, o número de artigos será de acordo com a quantidade de diferentes grandes áreas informadas.

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: (i) arquivo de indivíduos por palavras-chave (identificador do currículo, palavra-chave1,... palavra-chaveN); (ii) arquivo de palavras-chave por ano e grande área (palavra-chave, ano, frequência e grande área da publicação) e (iii) arquivo de frequência de palavras-chave por ano e quantidade de publicação (ano, frequência palavra-chave, quantidade de publicação e grande área da publicação).

Figura 4 – Exemplo de processamento dos métodos ISCooll original e adaptado



Fonte: Elaborado pelos autores

RESULTADOS

Os dados foram coletados em abril de 2017, totalizando 265.170 currículos de indivíduos com doutorado concluído das diversas áreas do conhecimento científico brasileiro. Dentre os currículos, cerca de 74% possuem data de última atualização nos últimos dois anos, o que comprova a contemporaneidade das informações. Para as análises foram considerados os dados de artigos publicados em anais de congressos e em periódicos referentes ao período de 1962 até 2016.

Neste ponto, para tornar as análises mais abrangentes e 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 em coautoria de mesma grande área do conhecimento associado ao conjunto união de suas respectivas palavras-chave. Assim, a tabela 2 apresenta o quantitativo de artigos e palavras-chave extraídos dos currículos dos doutores, considerando as visões geral e de colaboração.

Os dados apresentados correspondem a toda produção de artigos cadastrados nos currículos extraídos. A quantidade de artigos registrados nos currículos dos doutores representa cerca de 70% do número total de artigos contidos no repositório da Plataforma Lattes (DIAS, 2016), que, aliado à diversidade e contemporaneidade de atualização destes pondera a validade do conjunto analisado.

Portanto, é possível destacar considerável preferência dos doutores sobre artigos em congressos. Como esperado, o total de artigos na visão colaboração é cerca de 25% menor que o total da visão geral, devido aos artigos terem sido registrados em coautoria. Entretanto, o número total de palavras-chave da visão colaboração é maior que da visão geral, fato este devido aos autores cadastrarem diferentes palavras-chave e/ou grandes áreas para o mesmo artigo realizado em coautoria. Outra situação também esperada é quanto ao número de palavras-chave ser maior que o número de artigos, visto que cada artigo pode ter até 6 palavras-chave na Plataforma Lattes.

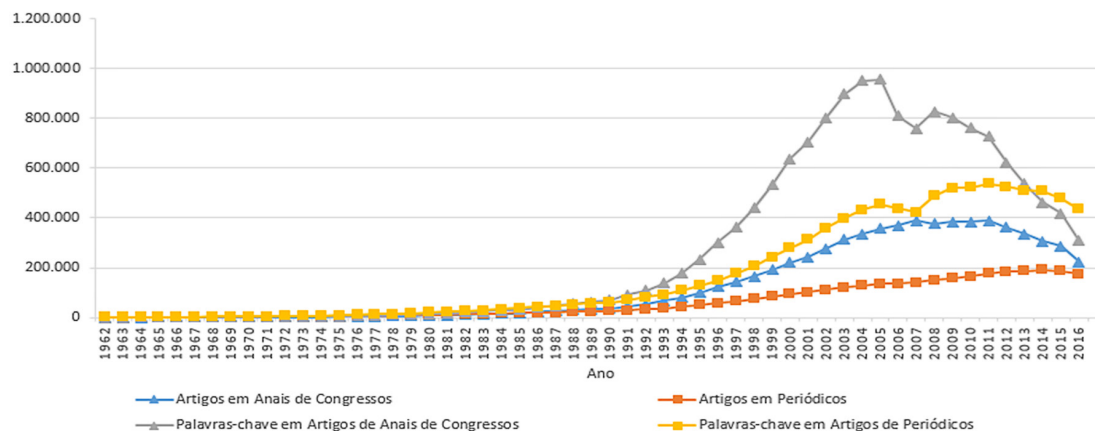
Para maiores detalhes, a figura 5, construída a partir da visão colaboração, ilustra os quantitativos dos arquivos e suas palavras-chave ao longo do período analisado.

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%)

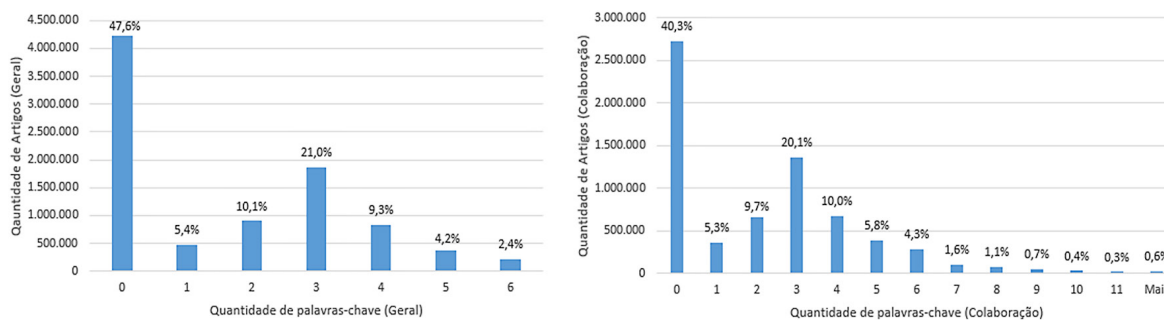
Fonte: Elaborado pelos autores. Dados da pesquisa

Figura 5 – Quantitativo de artigos e suas palavras-chave por ano



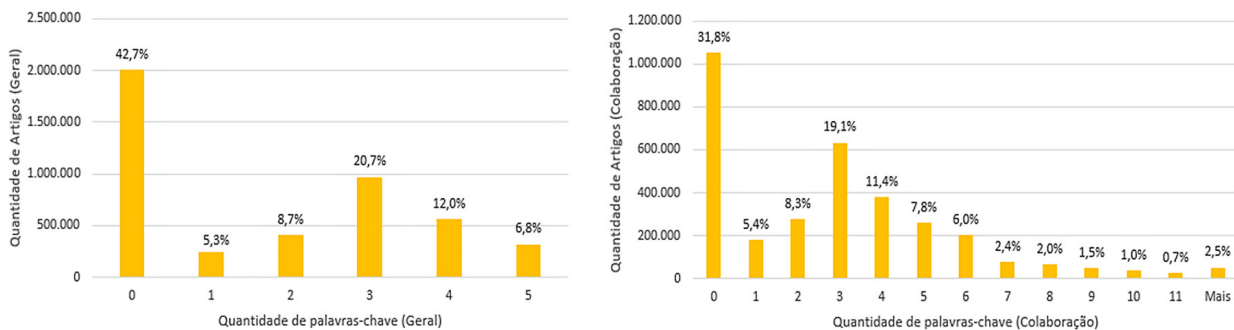
Fonte: Elaborado pelos autores

Figura 6 – Distribuição dos artigos de congressos pelo número de palavras-chave



Fonte: Elaborado pelos autores

Figura 7 – Distribuição dos artigos de periódicos pelo número de palavras-chave.



Fonte: Elaborado pelos autores.

Como pode ser observado, para a maioria dos anos as curvas das palavras-chave em anais de congressos e em periódicos apresentam o mesmo comportamento dos artigos. Outra situação é que o maior número de palavras-chave não ocorreu no mesmo ano do maior número de artigos. Percebe-se ainda que as palavras-chave tiveram crescimento considerável entre 1995 e 2005, pois em congressos cresceram 443% e os periódicos 197%. No entanto, a partir de 2005, houve queda de 67% em congressos e 18,9% em periódicos. A queda acentuada em congressos nos últimos anos demonstra maior interesse dos doutores quanto a periódicos, visto que a partir de 2014 o número de palavras-chave em periódicos superou congressos, fato este possivelmente devido à exigência da Capes considerar apenas artigos em periódicos em suas avaliações.

As figuras 7 e 8 apresentam os histogramas referentes a distribuição do número de artigos por quantidade de palavras-chave a partir das visões geral e colaboração respectivamente.

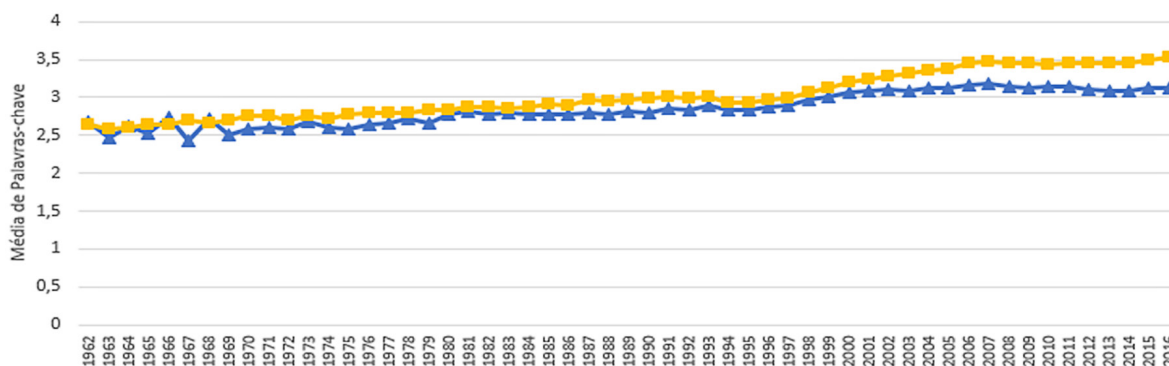
Destaca-se que grande parte dos artigos (em média 45% visão geral e 36% colaboração) não possui nenhuma palavra-chave. Além disso, cerca de 15% têm entre uma e duas palavras-chave. Adicionalmente, através da visão colaboração, é possível comprovar que os doutores que realizaram trabalhos em coautoria, inseriram diferentes palavras-chave para o mesmo artigo, sendo que os números de palavras são superiores à capacidade aceita pela Plataforma Lattes.

Assim, para possibilitar uma análise mais detalhada acerca da relação entre o número de palavras-chave utilizadas por artigo, a figura 8, com base na visão geral e considerando apenas o quantitativo dos artigos que possuem palavras-chave, apresenta a média de palavras-chave por artigos ao longo do período analisado.

Percebe-se que a média de palavras-chave por artigo em periódicos é superior a congressos para praticamente todo o período. E que a partir de 1999 a média de palavras-chave por artigo sempre foi superior a 3, destacando-se as médias em periódicos, que para o mesmo período cresceu cerca de 8% em relação a congressos.

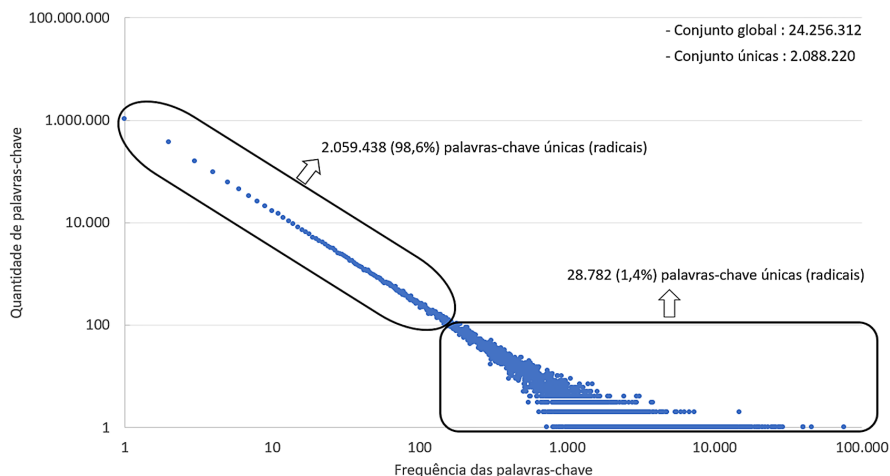
Uma informação muito importante para a compreensão sobre a organização do conjunto das palavras-chave em análise diz respeito ao quantitativo de palavras por frequência de utilização. Assim, com base na visão colaboração, a figura 9 apresenta a distribuição da quantidade de palavras-chave por frequência. Além disso, mostra também o quantitativo de palavras-chave únicas, ou seja, palavras que foram processadas pela etapa de limpeza e agrupamento dos dados.

Figura 8 – Média de palavras-chave por artigo publicado



Fonte: Elaborado pelos autores

Figura 9 – Distribuição de quantidade de palavras-chave por sua frequência



Fonte: Elaborado pelos autores.

Assim, observa-se que 24.256.312 palavras-chave (conjunto global), após processadas, se reduziram a 2.088.220 palavras-chave únicas. Dentre estas, apenas 1,4% das palavras-chave únicas possuem frequência superior a 100, enquanto 98,6% foram utilizadas menos de 100 vezes. Essa informação mostra que os doutores brasileiros têm concentrado mais esforços acerca de um conjunto específico de temas ao longo do tempo, e, com isso, confirmando uma das leis clássicas da Bibliometria, Lei de Zipf, na qual se descreve que um pequeno número de palavras é usado muito mais frequentemente.

A grande área do conhecimento é uma das informações requeridas durante o cadastramento dos artigos nos currículos da Plataforma Lattes. De acordo com Dias (2016), a lista das grandes áreas disponíveis na Plataforma segue, de modo geral, a mesma utilizada pela Capes, salvo pequenas distinções. Assim, torna-se factível entender quais são as contribuições científicas de cada uma para a ciência brasileira e, com isso, verificar se existem grandes áreas distintas trabalhando no mesmo tema de pesquisa. A seguir, a tabelas 3, construída com base nas visões colaboração, apresenta o quantitativo de artigos e suas palavras-chave por grande área do conhecimento.

Tabela 3 – Artigos em anais de congresso e periódicos e suas palavras-chave, visão colaboração

Grande Área	Anais de Congresso		Periódicos	
	Artigos	Palavras-chave	Artigos	Palavras-chave
Ciências Agrárias	1.055.864 (15,58%)	2.690.973 (18,13%)	391.365 (11,99%)	1.463.971 (15,45%)
Ciências Biológicas	1.052.861 (15,53%)	2.318.147 (15,62%)	487.447 (14,93%)	1.658.380 (17,61%)
Ciências da Saúde	1.459.741 (21,54%)	2.495.641 (16,81%)	813.826 (24,93%)	2.143.681 (22,76%)
Ciências Exatas e da Terra	899.138 (13,26%)	2.093.349 (14,10%)	439.256 (13,45%)	1.291.121 (13,71%)
Ciências Humanas	855.866 (12,62%)	1.845.203 (12,43%)	432.052 (13,23%)	1.106.906 (11,75%)
Ciências Sociais Aplicadas	422.929 (6,24%)	928.006 (6,25%)	291.937 (8,94%)	665.511 (7,06%)
Engenharias	696.512 (10,27%)	1.916.713 (12,91%)	210.263 (6,44%)	709.994 (7,53%)
Linguística, Letras e Artes	228.704 (3,37%)	465.312 (3,13%)	136.480 (4,18%)	320.626 (3,40%)
Outra	105.056 (1,55%)	85.884 (0,57%)	61.367 (1,88%)	56.894 (0,60%)
Total	6.776.671 (100%)	14.839.228 (100%)	3.263.993 (100%)	9.414.084 (100%)

Fonte: Elaborado pelos autores. Dados da pesquisa

Logo, em relação às palavras-chave em anais de congressos, destaca-se a ciências agrárias, pois representa mais de 18% do conjunto total de palavras-chave, contudo não é a responsável pelo maior número de artigos publicado. Isto aconteceu devido os indivíduos que atuam na grande área ciências da saúde (responsável pelo maior número de artigos publicados) não terem associados palavras-chave em parte dos artigos cadastrados em seus currículos. A respeito dos periódicos, cerca de 26% das palavras-chave referem-se à produção da grande área de ciências da saúde, que nesse caso também foi responsável pelo maior número de artigos publicados. Com base nos baixos valores encontrados da grande área Outra, optou-se pela desconsideração das informações referentes a esta área para o restante das análises.

Em continuação, para conhecer os principais tópicos de pesquisas de cada grande área do conhecimento, todo o conjunto de palavras-chave únicas, resultado do processamento dos dados com base na visão colaboração, foi agrupado de acordo com a área associada ao artigo, onde, posteriormente, foram ranqueadas de acordo com a medida de importância de tópicos baseada em frequência. Assim, a tabela 4 apresenta o ranqueamento das principais palavras-chave utilizadas nos artigos científicos publicados entre 1962 e 2016 em cada grande área do conhecimento.

Destaca-se a palavra-chave educação, utilizada 50.333 vezes pelos doutores que publicaram na grande área ciências humanas, e concomitantemente, a palavra mais frequente dentre todas as grandes áreas, tendo aparecido, 12.693 a mais que enfermagem, palavra-chave mais utilizada em publicações da grande área ciências da saúde. Isso mostra que a grande área ciências humanas contribuiu cerca de 66% para tornar a palavra-chave educação a mais utilizada pelos doutores brasileiros em toda história, uma vez que foi utilizada 75.855 vezes no total. Além disso, é a palavra-chave que mais vezes apareceu entre as principais de todas as grandes áreas, ou seja, também é tema de pesquisa de interesses nas ciências da saúde, ciências exatas e da terra e ciências sociais aplicadas.

Por fim, outras palavras-chave são utilizadas por diferentes grandes áreas. Para facilitar visualização desta relação, a tabela 5 apresenta o coeficiente de similaridade das principais palavras-chave entre cada par de grande área do conhecimento. Neste contexto, coeficiente de similaridade mede o quão dois conjuntos principais de palavras-chave são iguais (a taxa da interseção entre eles), independentemente da ordem do ranqueamento destas palavra-chave em cada conjunto.

Tabela 4 – As palavras-chave mais frequentes em cada grande área entre 1962 e 2016

Ranque	Ciências Agrárias	Frequência	Ciências Biológicas	Frequência	Ciências da Saúde	Frequência
1	Bovinos	22.570	Taxonomia	15.110	Enfermagem	37.640
2	Cães	18.890	Ratos	11.770	Epidemiologia	26.527
3	Sementes	13.177	Trypanosoma Cruzi	10.912	Idoso	20.819
4	Soja	13.049	Amazônia	9.788	Crianças	20.470
5	Milho	12.984	Ecologia	9.105	Diagnóstico	13.365
6	Ovinos	12.417	Cerrado	8.977	Adolescentes	13.098
7	Controle Biológico	11.532	Morfologia	8.504	Qualidade de Vida	11.403
8	Qualidade	11.167	Biodiversidade	8.255	Obesidade	11.290
9	Equinos	10.508	Mata Atlântica	7.588	Ratos	10.818
10	Irrigação	10.342	Peixes	7.511	Educação Física	10.661
11	Suínos	9.633	Conservação	7.201	Tratamento	10.567
12	Nutrição	9.566	Citogenética	6.608	Aids	9.908
13	Cultivo	9.443	Anatomia	6.329	Educação e Saúde	9.703
14	Germinação	9.429	Plantas Medicinais	6.291	Câncer	8.639
15	Adução	8.810	Epidemiologia	5.647	Educação	8.494

(Continua)

Tabela 4 – As palavras-chave mais frequentes em cada grande área entre 1962 e 2016

(Conclusão)

Ranque	Ciências Exatas e da Terra	Frequência	Ciências Humanas	Frequência	Ciências Sociais Aplicadas	Frequência
1	Sensoriamento Remoto	7.885	Educação	50.333	Comunicação	9.974
2	Geoprocessamento	5.378	Formação do Professor	37.174	Políticas Públicas	7.825
3	Biodiesel	4.426	Políticas Públicas	14.583	Estratégia	6.597
4	Ensino de Química	4.230	Gênero	12.700	Sustentabilidade	6.522
5	Amazônia	4.136	História	12.606	Educação	5.965
6	Geoquímica	4.015	Currículo	11.595	Inovação	5.660
7	Catalisador	3.956	Educação Ambiental	11.418	Jornalismo	5.308
8	Sedimentos	3.685	Cultura	11.001	Turismo	5.216
9	Adsorção	3.557	Memória	10.386	Marketing	5.073
10	Síntese	3.475	Trabalho	10.265	Design	5.033
11	Educação e Matemática	3.397	Psicanálise	9.192	Brasil	4.800
12	Ensino de Física	3.384	Identidade	9.098	Desenvolvimento Sustentável	4.556
13	Educação	3.200	Educação Infantil	8.987	Desenvolvimento Regional	4.415
14	Espectroscopia	2.999	História da Educação	8.777	Meio Ambiente	4.297
15	Flavonoides	2.938	Educação Especial	8.620	Administração	4.279

Ranque	Engenharias	Frequência	Linguística, Letras e Artes	Frequência
1	Compósitos	6.088	Literatura	6.350
2	Otimização	5.535	Leitura	6.024
3	Concreto	5.300	Análise Discursiva	5.893
4	Simulação	4.994	Discurso	5.650
5	Corrosão	4.937	Literatura Brasileira	5.612
6	Reciclagem	4.388	Identidade	4.587
7	Biodiesel	4.360	Formação do Professor	4.075
8	Modelagem	4.155	Ensino	4.019
9	Ergonomia	3.977	Memória	3.864
10	Biomateriais	3.958	Artes	3.663
11	Cerâmicas	3.852	Poesia	3.592
12	Propriedades Mecânicas	3.755	Tradução	3.237
13	Elementos Finitos	3.715	Literatura Comparada	3.128
14	Resíduos Sólidos	3.611	Língua Portuguesa	3.094
15	Adsorção	3.579	Linguagem	3.083

Fonte: Elaborado pelos autores. Dados da pesquisa.

Tabela 5 – Coeficiente de similaridade por grandes áreas entre 1962 e 2016

	Ciências Agrárias	Ciências Biológicas	Ciências da Saúde	Ciências Exatas e da Terra	Ciências Humanas	Ciências Sociais Aplicadas	Engenharias	Linguística, Letras e Artes
Ciências Agrárias	X	0%	0%	0%	0%	0%	0%	0%
Ciências Biológicas		X	13,33%	6,66%	0%	0%	0%	0%
Ciências da Saúde			X	6,66%	6,66%	6,66%	0%	0%
Ciências Exatas e da Terra				X	6,66%	6,66%	13,33%	0%
Ciências Humanas					X	13,33%	0%	20,00%
Ciências Sociais Aplicadas						X	0%	0%
Engenharias							X	0%
Linguística, Letras e Artes								X

Fonte: Elaborado pelos autores. Dados da pesquisa.

verifica-se que apenas ciências agrárias não possui relação de similaridade dentre seus respectivos principais tópicos com os assuntos centrais das outras grandes áreas. em contrapartida, todas as grandes áreas restantes se relacionam entre si quanto ao interesse em pelo menos um tópico de pesquisa. Dentre estas, destaca-se a ciências exatas e da terra, que possui o maior número de relações: ciências biológicas (amazônia); ciências da saúde (educação); ciências humanas (educação); ciências sociais aplicadas (educação) e; engenharias (adsorção, biodiesel). No entanto, quando observada a força do relacionamento, evidenciam-se ciências humanas e linguística, letras e artes, pois possuem a maior semelhança (20%) entre os principais tópicos de pesquisas que trabalham. São eles: identidade, formação do professor e memória.

CONSIDERAÇÕES FINAIS

Tendo em vista o estudo realizado, ressalta-se que este tipo de análise se caracteriza como importante mecanismo, pois possibilita identificar quais os tópicos mais impactantes dentre uma comunidade de pesquisa. Ao analisar as palavras-chave dos currículos cadastrados na Plataforma Lattes, é possível considerar publicações realizadas em anais de congressos e periódicos nacionais não indexados, o que não é factível quando verificado repositório de dados internacionais. Logo, é possível obter uma visão mais precisa dos tópicos mais investigados pelos pesquisadores brasileiros.

Através dos resultados encontrados, foi possível apresentar uma caracterização geral sobre as palavras-chave utilizadas pelos doutores brasileiros. Com base na medida de frequência, nota-se que pequeno número de palavras-chave recebe maior atenção por parte dos doutores, enquanto grande quantidade de assuntos foram poucos estudados. No âmbito das grandes áreas do conhecimento, tem-se ciências agrárias como a responsável pelo maior número de palavras-chave em congressos (cerca de 18%) e ciências da saúde em periódicos (cerca de 23%).

Em seguida, observou-se que linguística, letras e artes e ciências humanas possuem maior interesse em comum, tendo semelhança em 20% dos principais tópicos estudados. Já ciências agrárias não possui relação com nenhuma outra grande área.

Entretanto, apesar da validade dos resultados, percebe-se que a análise das palavras-chave utilizando apenas medida de importância baseado em frequência sofre influência quantitativa, que no caso da Plataforma Lattes, que possui artigos publicados nas mais variadas conferências e periódicos com diferentes níveis de qualidades, análises apenas quantitativas podem levar a percepções equivocadas. Com isso, em trabalhos futuros, espera-se que sejam realizadas análises que consideram fatores temporais para mapear os principais tópicos de pesquisas de cada grande área ao longo do tempo, análises baseadas em técnicas de redes sociais para determinar as palavras-chave mais centrais, e ainda, características qualitativas apropriadas ao contexto da Plataforma Lattes. Além disso, devido à pouca padronização existente no registro das palavras-chave por parte dos pesquisadores quanto às áreas do conhecimento, espera-se também a utilização de um thesaurus para tal finalidade.

REFERÊNCIAS

BORGES, V. A.; NOGUEIRA, B.M.; BARBOSA, E. F. Uma análise exploratória de tópicos de pesquisa emergentes em Informática na Educação. *Revista Brasileira de Informática na Educação*, v. 23, n. 1, p. 1-13, 2015. DOI: <http://dx.doi.org/10.5753/rbie.2015.23.01.85>

BRITO, A. G. C.; QUONIAM, L.; MENA-CHALCO, J. P. Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação*, v.28, n.1, p. 77-86, 2016.

CARNEIRO, D.M. C&T em prol da cidadania. Fapemig. 2003. Disponível em: http://lqes.iqm.unicamp.br/canal_cientifico/pontos_vista/pontos_vista_artigos_opiniao27-1.html.

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>

CHEN, G. et al. Identifying the research focus of Library and Information Science institutions in China with institutions keywords. *Scientometrics*, v. 103, n. 2, p. 707-724, 2015. Disponível em: <http://link.springer.com/article/10.1007/s11192-015-1545-2>.

CHOI, J.; SANGYOON, Y.; LEE, K.C. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management*, v. 48, n. 8, p. 371-381, 2011. DOI: <http://dx.doi.org/10.1016/j.im.2011.09.004>

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. Disponível em <http://seer.ufrgs.br/index.php/EmQuestao/article/view/53259>

DIAS, T. M. R. Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes. 2016. 181p. Tese (Doutorado em Modelagem Matemática e Computacional) - Programa em Pós-Graduação em Modelagem Matemática e Computacional, CEFET-MG, Belo Horizonte, 2016.

DIGIAMPIETRI, L. A. Análise da rede social acadêmica brasileira. 2015. 160 p. Tese (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. In: ENCONTRO NACIONAL DE INTELIGÊNCIA ARTIFICIAL E COMPUTACIONAL - ENIAC , 11., 2014, São Paulo. *Anais...* São Paulo, 2014. p. 1-6.

HERADIO, R. et al. A bibliometric analysis of 20 years of research on software product lines. *Information and Software Technology*, v. 72, p. 1-15, 2016. DOI: <http://dx.doi.org/10.1016/j.infsof.2015.11.004>

HONG, Y. et al. Knowledge structure and theme trends analysis on general practitioner research: A Co-word perspective. *Bio Med Central*, v. 17, 2016. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4734860>

KHAN, G. F.; WOOD, J. Information technology management domain: emerging themes and keyword analysis. *Scientometrics*, v. 105, n. 2, p. 959-972, 2015. 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.

MCCLOSKEY, D. N. The rhetoric of economics. University of Wisconsin. 1998.

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, Montreal, Canadá, 2013.

- MRYGLOD, O. et al. Quantifying the evolution of a scientific topic: reaction of the academic community to the Chernobyl disaster. *Scientometrics*, v. 106, p. 1151- 1166, 2016. Disponível em: <<http://link.springer.com/article/10.1007/s11192-015-1820-2>>.
- PEREIRA, J.C.R. et al. Who's who and what's what in Brazilian Public Health Sciences. *Scientometrics*, v. 73, p. 37- 52, 2007. DOI: <http://dx.doi.org/10.1007/s11192-007-1787-8>
- PETER, R.S. et al. Epidemiologic research topics in Germany: a keyword network analysis of 2014 DGEpi conference presentations. *European Journal of Epidemiology*, v. 31, p. 635- 638, 2016. Disponível em: <<https://link.springer.com/article/10.1007%2Fs10654-016-0141-y>>.
- 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. DOI: <http://dx.doi.org/10.1016/j.ijproman.2014.04.011>
- RAVIKUMAR S.; AGRAHARI A.; SINGH S.N. Mapping the intellectual structure of scientometrics: a co-word analysis of the journal *Scientometrics* (2005-2010). *Scientometrics*, v. 102, n. 1, p. 929-955, 2015. Disponível em: <<http://link.springer.com/article/10.1007/s11192-014-1402-8>>
- RONDA-PUPO, G. Knowledge map of Latin American research on management: Trends and future advancement. *SAGE - Social Science Information*, v. 55, n. 1, p. 3 - 27, 2015. Disponível em: <<http://journals.sagepub.com/doi/abs/10.1177/0539018415610225>>.
- 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. Dissertação (Doutorado) – Programa em Pós-Graduação em Saúde Pública, Universidade de São Paulo, São Paulo, 2005.
- SILVA, F.N. et al. Using network science and text analytics to produce surveys in a scientific topic. *Journal of Informetrics*, v. 10, n. 2, p. 487 - 502, 2016. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S1751157715301966>>.
- 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>>.
- 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.
- TRUCOLO, C.C. Análise de tendências em redes sociais acadêmicas. 2016. Dissertação (Mestrado). Programa em Pós-Graduação em Sistemas de Informação, São Paulo, 2016.
- VINKERS, C. H.; TIJDINK, J. K.; OTTE, W. M. J. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ*, 2015. Disponível em: <<http://www.bmj.com/content/351/bmj.h6467>>.
- ZHANG, K. et al. A bibliometric analysis of research on carbon tax from 1989 to 2014. *Renewable and Sustainable Energy Reviews*, v. 58, p. 297-310, 2016. DOI: <http://dx.doi.org/10.1016/j.rser.2015.12.089>
- 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. DOI: <http://dx.doi.org/10.1007/s11192-013-1019-3>
- YI, S., CHOI, J. The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics*, v. 90, n. 3, p. 1016-1026, 2012. DOI: <http://dx.doi.org/10.1007/s11192-011-0560-1>