



UMA ANÁLISE TEMPORAL DOS PRINCIPAIS TÓPICOS DE PESQUISAS EM ENGENHARIAS NO BRASIL

Jether Gomes

jethergomes@yahoo.com.br

Thiago M. R. Dias

thiago@div.cefetmg.br

Gray F. Moita

gray@dppg.cefetmg.br

Tales H. J. Moreira

tales.info@gmail.com

Programa de Pós-graduação em Modelagem Matemática e Computacional

Centro Federal de Educação Tecnológica de Minas Gerais

Av. Amazonas, 7576, Nova Gameleira, 30510-000, Belo Horizonte, MG, Brasil.

Resumo. A internet e seus serviços são fatores fundamentais que impulsionaram o significativo crescimento da divulgação de artigos científicos nas últimas décadas. Consequentemente, verifica-se um aumento global de pesquisadores de todas as áreas do conhecimento interessados em compreender o desenvolvimento científico. Uma vez que, essa compreensão pode fornecer insumos importantíssimos para auxiliar à diversos tipos de tomada de decisão, como, por exemplo, auxiliar pesquisadores na escolha de novos temas de pesquisas, tornar grupo de pesquisadores mais produtivos, e ainda, servir de base para a construção de políticas científicas visando novos avanços na ciência. Sendo assim, este trabalho tem como objetivo identificar e analisar temporalmente os principais tópicos investigados entre 1962 e 2016 pelos doutores brasileiros que atuam nas áreas de engenharias. Para isso, são realizados estudos bibliométricos sobre todo o conjunto de palavras-chave dos artigos científicos referentes à grande área engenharias, publicados em anais de congressos e em periódicos pelos doutores que possuem currículos cadastrados na Plataforma Lattes.

Palavras-chave: Palavras-chave, Tópicos de Pesquisas, Bibliometria, Plataforma Lattes.

1. INTRODUÇÃO

A internet e seus serviços foram fundamentais para o crescimento impressionante do número de publicações de artigos científicos nas últimas décadas. Segundo Dias (2016), a grande facilidade para acessar os artigos disponibilizados na internet contribui diretamente para a ampliação do conhecimento. Por outro lado, a divulgação deste conhecimento tem ajudado a compreender melhor o mundo, contribuindo com o desenvolvimento da sociedade (Ferreira, 2010).

A implementação do conhecimento científico é tarefa indispensável para o desenvolvimento econômico e social (Saes, 2005). No entanto, muitas das vezes, a implementação do conhecimento não é uma tarefa trivial, visto que, geralmente existe uma quantia limitada de recursos para fomentar pesquisas e um grande número de pesquisadores ou instituições interessadas nesses recursos (Digiampietri, 2015). Diante disto, tem-se cada vez mais necessidade de racionalidade e objetividade na aplicação dos poucos recursos disponíveis.

Para Saes (2005), tão importante quanto ter consideráveis investimentos em pesquisas, é ter habilidade para controlar, compreender e medir o patamar científico das nações e/ou grupos individualizados. Logo, torna-se indispensável estudar, para compreender e medir, a produção científica para auxiliar na decisão das prioridades científicas de modo que alcance os pressupostos de racionalidade e objetividade.

Segundo Yi e Choi (2012), a compreensão sobre a evolução da produção científica pode promover novos avanços na ciência. 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, tecnologia e inovação, possibilitando gerar parâmetros para nortear esforços e investimentos com intuito de impulsionar resultados de pesquisa.

Em consequência disto, verifica-se um crescimento global quanto ao número de pesquisadores de todas as áreas para extrair conhecimentos acerca do que tem sido desenvolvido em ciência (Dias 2016; Digiampietri 2015; Mena-Chalco et. al. 2014; Dias et. al. 2014). Neste ponto, é importante ressaltar que os trabalhos podem ser realizados sob diferentes perspectivas, como: análise de citações, indicadores de produtividade, colaboração científica e análise de tópicos de pesquisa. Neste último caso, um tópico pode ser entendido como um termo (ou descritor) que representa um dos assuntos associados a um determinado documento (Borges et. al., 2015).

Entretanto, no âmbito nacional, pouquíssimos são os trabalhos que realizam análise tópicos de pesquisas, e, quando acontecem, geralmente utilizam uma quantidade restrita de dados que são oriundas de repositórios científicos internacionais específicos, e, por isso, não podem representar o que é produzido no Brasil. Assim, analisar dados de repositórios que englobe diversos tipos de publicações, principalmente em veículos nacionais e de diversas áreas do conhecimento, passa a ser uma tarefa relevante para a compreensão sobre a evolução da ciência brasileira. Neste contexto, Ferraz et. al. (2014) destaca que o grande volume de dados existente no repositório curricular da Plataforma Lattes representa o instrumento de maior importância para estudos sobre a produção científica brasileira.

Uma das abordagens mais interessantes para estudos de tópicos de pesquisas é a análise das palavras-chave de um conjunto de publicações científicas, pois, são inseridas cuidadosamente por seus respectivos autores e fornecem uma possibilidade de descrever os assuntos principais que permeiam o trabalho (McCloskey, 1998; Yi e Choi, 2012; Khan e Wood,

2015). Com isso, no atual trabalho, palavras-chave de artigos científicos são igualmente referenciadas como tópicos de pesquisas.

Em continuação, para Pritchard (1969) a bibliometria se destaca como uma das principais ciências métricas de análise de conteúdo, podendo ser aplicada a repositórios de dados científicos com o intuito de se obter informações quantitativas sobre publicações. Enquanto em Dias (2016), o autor destaca que com a utilização da bibliometria é possível identificar as tendências e o crescimento do conhecimento científico em diversas áreas, observar a dispersão do conhecimento científico, auxiliar políticas para investimentos e entender como acontece a evolução científica.

Assim sendo, este trabalho tem como objetivo realizar uma análise temporal sobre todas as palavras-chave de artigos científicos referente a grande área engenharias, publicados em anais de congressos e em periódicos entre 1962 e 2016 pelos doutores brasileiros que possuem currículos cadastrados na Plataforma Lattes. O estudo contempla análises bibliométricas sobre as palavras-chave em cada quinquênio ao longo do tempo, afim de identificar e mapear os principais tópicos de investigação nas engenharias pelos pesquisadores doutores brasileiros.

2. TRABALHOS RELACIONADOS

Os trabalhos que identificam e analisam tópicos de pesquisa geralmente se baseiam em estudos relacionados a termos extraídos dos títulos, resumos ou em palavras-chaves de publicações científicas (Khan e Wood, 2015; Choi et. al., 2011). Por exemplo, Kauer e Moreira (2013) aplicaram o método de mineração de textos como forma de identificar a evolução dos tópicos abordados no Simpósio Brasileiro de Banco de Dados (SBBD) ao longo dos anos. A análise foi realizada sobre os termos existentes nos 475 resumos de artigos publicados entre 1989 e 2012. Este conjunto de documentos foi particionado em 23 intervalos de tempo, onde cada intervalo representa uma edição do SBBD. Os resultados mostraram que nos anos iniciais, os temas eram mais diferentes entre si a cada edição.

Em Souza et. al. (2014), foram analisadas redes semânticas, construídas a partir de termos extraídos dos 14.406 títulos de chamados de um sistema de atendimento eletrônico de uma universidade, através de técnicas de redes sociais. Os resultados encontrados a partir de análises das métricas de centralidade permitiram visualizar quais os problemas mais recorrentes, a fim de facilitar a tomada de decisão e a definição de tendências nas solicitações de determinados usuários e/ou setores.

Com a mesma ideia, e um novo trabalho, Souza et. al. (2015) analisaram uma rede semântica criada a partir dos termos presentes nos títulos dos artigos publicados na área de Sistemas Embarcados para destacar os temas mais relevantes. Os dados contabilizam 11.303 artigos entre os anos de 1981 e 2014 cadastrados na *IEEE Explorer*. Para encontrar os termos em destaques, aplicaram métricas de redes sociais baseadas em medidas de centralidade. Em seguida, realizaram uma análise temporal dos termos mais relevantes e aplicaram uma regressão linear para gerar a equação de previsão de tendências para os próximos cinco anos.

O trabalho de Lucas e Lara (2014) teve uma abordagem baseada em palavras-chave utilizadas em artigos científicos, como forma de identificar quais são as mais representativas num determinado corpus. O corpus em estudo foi composto por 354 artigos publicados entre 1997 e 2011 referente ao tema *Knowledge Organization* extraídos dos repositórios de dados da *Library Information Science Abstract*, *Library, Information Science & Technology Abstracts*,

Scopus e *Web of Science*. Após as análises realizadas, foram identificadas 927 palavras-chave que melhor descrevem o corpus. Além disso, foram evidenciadas as relações entre elas como forma de viabilizar a percepção das relações entre autores com alta representatividade e autores pouco produtivos. Assim, foi apresentado como autores pouco produtivos são influenciados pelas autoridades de determinada área.

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 ao longo dos anos. Segundo eles, entender o comportamento de utilização desses termos pode gerar insumos para auxiliar na interpretação sobre o desenvolvimento da ciência. Para tanto, definiram conjuntos de termos para cada classe, sendo: 25 selecionados por especialistas e 100 aleatoriamente. Os termos foram quantificados individualmente por período de tempo e comparados com os livros publicados entre 1975 e 2009 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 a evolução da literatura científica. Contudo, os autores listaram diversos fatores que poderiam estar contribuindo para o exagero dos pesquisadores na utilização de uma linguagem positiva perante seus resultados.

Em sua pesquisa de doutorado, Saes (2005) analisou a produção científica brasileira no campo da saúde. A análise incluiu 3.066 revistas desse campo do conhecimento, correspondendo a um total de 38.349 artigos publicados entre 1990 e 2002 extraídos da *Web of Knowledge*. Os métodos bibliométricos e a co-ocorrência de termos foram os instrumentos utilizados para medir a atividade científica. Os termos que representam os descritores e identificadores dos artigos no repositório de dados foram relacionados com outras variáveis, como autores, ano de publicação e filiação da publicação. A partir daí, foi possível identificar redes de cooperação entre pesquisadores e instituições, bem como identificação de clusters entre autores e os termos descritores do campo do conhecimento estudado.

Por fim, Hong et. al. (2016) realizaram a combinação de análises bibliométricas, co-ocorrência 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 foram coletados no repositório da *PubMed*, sendo 10.704 publicações e seus respectivos 42.695 termos *Mesh* associados. Estes dados foram agrupados e analisados de acordo com as respectivas faixas de tempo: 1999 a 2004, 2004 a 2008 e 2009 a 2014, para facilitar a verificação da evolução dos tópicos estudados. Além dos resultados fornecerem uma visão clara acerca do assunto, também apontaram que temas relacionados a idosos serão amplamente discutidos no futuro, devido ao crescimento da população idosa. Entretanto, os autores citam duas limitações: (1) os dados analisados podem não representar em totalidade a atuação dos médicos clínicos gerais e (2) apenas os termos mais frequentemente foram analisados, sendo possível não mostrar algum termo emergente.

3. DESENVOLVIMENTO

O repositório de dados utilizado para realização deste trabalho foi a Plataforma Lattes do Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq). Esta plataforma tem como objetivo integrar os sistemas de informação dos órgãos federais brasileiros, otimizando o processo de gestão de Ciência e Tecnologia (C&T) (Dias et al., 2014).

A escolha do repositório da Plataforma Lattes está relacionada ao fato de tratar da integração de dados acadêmicos e profissionais de mais de 5 milhões de indivíduos de todas as áreas do conhecimento (Dias, 2016). Esses dados estão disponíveis livremente na internet, porém, ainda não foram amplamente analisados (Digiampietri, 2015). O repositório da Plataforma Lattes também é utilizado como fonte de informação para órgãos que avaliam o Sistema Nacional de Pós-graduação do Brasil, e para as agências de fomento e financiamento de pesquisas e/ou ofertas de bolsas.

Apesar dos dados disponibilizados na internet, estes são apenas visualizados através de uma interface de consulta que apresenta cada currículo individualmente, não permitindo uma análise mais ampla. Sendo assim, para uma análise mais abrangente sobre grupos, instituições ou até de todo o conjunto de cientistas brasileiros, técnicas e ferramentas para a extração e análises desses dados científicos se fazem necessárias (Dias, 2016).

A extração dos currículos dos doutores na versão *XML* (*eXtensible Markup Language*) foi realizada através do *LattesDataXplorer*, proposto e desenvolvido por Dias (2016) para coletar os dados curriculares do repositório da Plataforma Lattes. A Figura 1 apresenta uma visão geral sobre o processo de extração da ferramenta *LattesDataXplorer*.

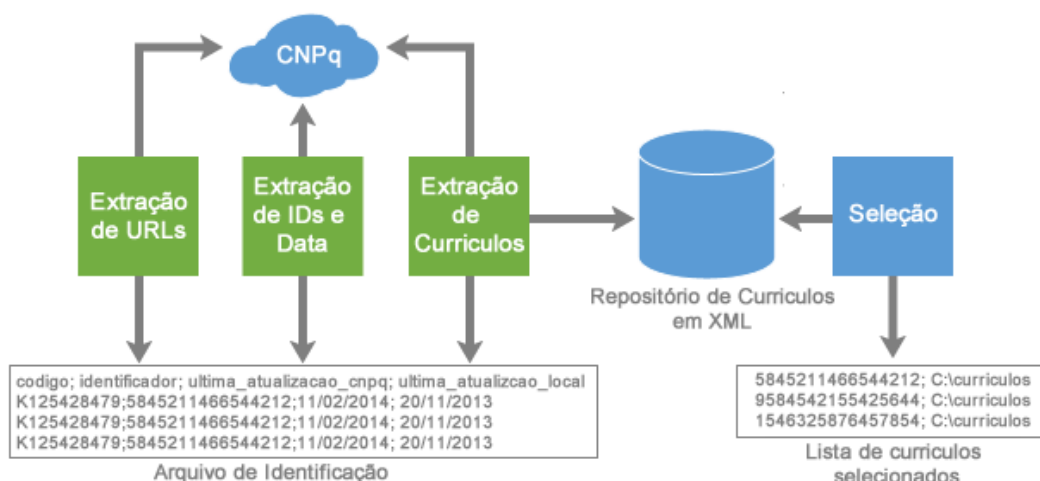


Figura 1. Processo de extração de dados *LattesDataXplorer* (Dias, 2016).

O processo de extração dos currículos foi realizado em três etapas: (1) extração de *URLs*, responsável por adquirir as referências únicas para os currículos cadastrados, e assim possibilitar o acesso individual a cada currículo, (2) extração de *Ids* e *Data*, para possibilitar o acesso a cada currículo e extrair seu identificador, bem como a data da última atualização, e (3) extração dos currículos, responsável pelo download e armazenamento em disco dos arquivos *XML* cuja data de atualização na Plataforma Lattes seja divergente da data de atualização do currículo armazenado localmente (DIAS, 2016).

Em seguida, foi desenvolvido um arcabouço de componentes para suportar as análises desejadas (Figura 2), onde os componentes “filtragem” e “tratamento” são responsáveis pelo processo de seleção, tratamento e modelagem dos dados dos currículos que são fundamentais para as análises da atual pesquisa, e, com isso, diminuir o tempo de processamento computacional.

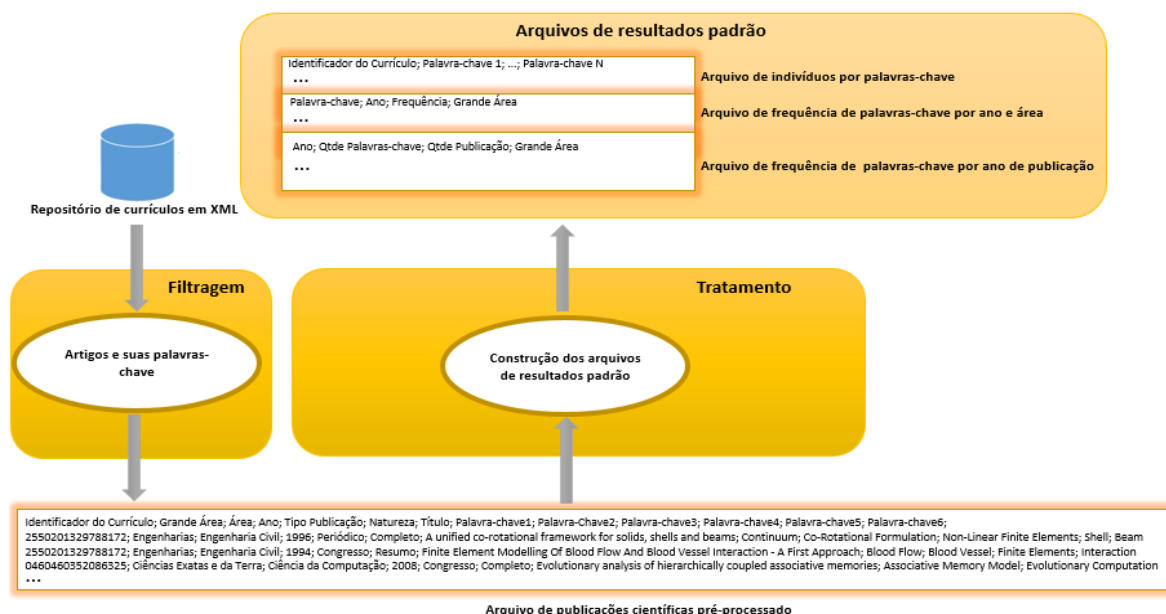


Figura 2. Arcabouço de componentes desenvolvidos para filtragem e tratamento dos dados.

O componente de “filtragem” realiza a etapa de mineração nos currículos para extrair as informações dos artigos, armazenando-as à parte num arquivo de publicações científicas, com isso, definindo o conjunto de dados centrais a serem estudados. As informações dos artigos incluem: identificador do currículo; grande área da publicação; ano de publicação; tipo de publicação; título e palavras-chave. Para este trabalho, foi considerado todas as publicações em anais de congressos e em periódicos que possuem como grande área, as engenharias.

Por outro lado, o componente de “tratamento”, processa os dados do arquivo de publicações científicas para tratá-los e caracterizá-los, e, a partir destes, constrói um conjunto dos arquivos de resultados padrão para facilitar as análises. Esse componente realiza basicamente três etapas, a saber: *limpeza e agrupamento dos dados*, *normalização dos dados* e *construção dos arquivos*.

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

Na etapa de *normalização dos dados* são processados todos os artigos do arquivo de publicações científicas para identificar e tratar os trabalhos realizados em colaboração. Esse procedimento é realizado para dados de um mesmo artigo científico não ser considerado mais de uma vez durante as análises. Para tanto, fez-se necessário uma adaptação no método *ISCool* proposto por Dias e Moita (2015) para identificar colaborações científicas em grande volume

de dados com custo linear. Assim, enquanto o método *ISCool* original utiliza um dicionário para vincular os artigos (chaves de um dicionário) a seus autores (identificadores dos currículos), o método *ISCool* adaptado adota um dicionário para vincular os títulos dos artigos ao conjunto união das respectivas palavras-chave.

A *construção 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 trabalhos realizados em 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) e (3) *Arquivo de frequência de palavras-chave por ano e quantidade de publicação* (Ano, frequência palavra-chave, quantidade de publicação e grande área).

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)*. 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, totalizando 265.170 currículos de doutores, onde 23.437 destes currículos têm as engenharias como principal grande área de atuação. Para visualizar a contemporaneidade do conjunto de dados a serem analisados, a Figura 3 apresenta a distribuição dos currículos dos doutores que atuam nas engenharias com base na última data de atualização.

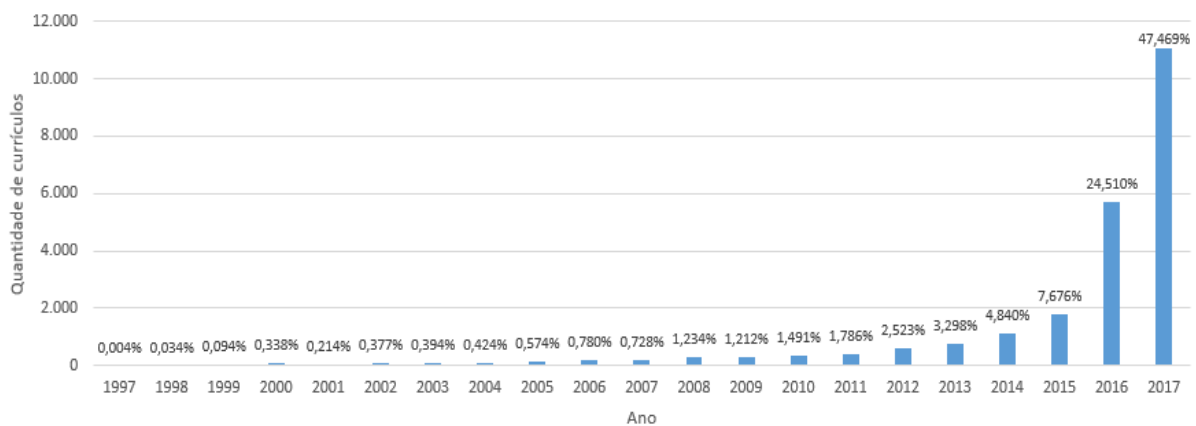


Figura 3. Distribuição dos currículos pela data da última atualização.

Como pode ser observado, cerca de 72% dos currículos possuem data de atualização recente, mais especificamente em 2016 (24,510%) e 2017 (47,469%). Apesar disto, vale ressaltar que, de acordo com Dias (2016), a não atualização do restante dos currículos pode ter motivos variados e de difícil reconhecimento.

Para as análises foram considerados todos os artigos publicados pelos doutores em anais de congressos e em periódicos entre 1962 e 2016 referentes a grande área engenharias. Neste ponto, para 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 1 apresenta o quantitativo dos dados de artigos extraídos dos currículos dos doutores que contribuíram com a grande área engenharias.

Tabela 1. Quantitativo de artigos científicos e suas palavras-chave.

	Tipo de Publicação	Artigos	Palavras-chave vinculadas aos artigos
Geral	Anais de Congresso	930.790 (75,4%)	1.821.212 (73,8%)
	Periódicos	304.097 (24,6%)	644.829 (26,2%)
	Total	1.234.887 (100%)	2.466.041 (100%)
Colaboração	Anais de Congresso	696.512 (76,8%)	1.916.713 (72,9%)
	Periódicos	210.263 (23,2%)	709.994 (27,1%)
	Total	906.775 (100%)	2.626.707 (100%)

Percebe-se que os doutores têm dedicado mais de 70% dos esforços em relação as publicações e preenchimento de palavras-chave em anais de congressos para ambas visões consideradas. Como esperado, também pode ser notado que o número de publicações da visão colaboração é inferior a visão geral, fato este, devido a existência de artigos realizados em coautoria cadastrados em diferentes currículos. Por outro lado, o número de palavras-chave na visão colaboração é superior ao da visão geral, devido os doutores terem associados diferentes palavras-chave para o mesmo artigo publicado em conjunto.

Entretanto, para as próximas análises, apenas as palavras-chave referente a visão colaboração são consideradas. Assim, com intuito de compreender sobre a distribuição de todo o conjunto das palavras-chave quanto a frequência de utilização, a Figura 4 apresenta a quantidade de palavras-chave pela respectiva 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, conforme descrito na seção anterior.

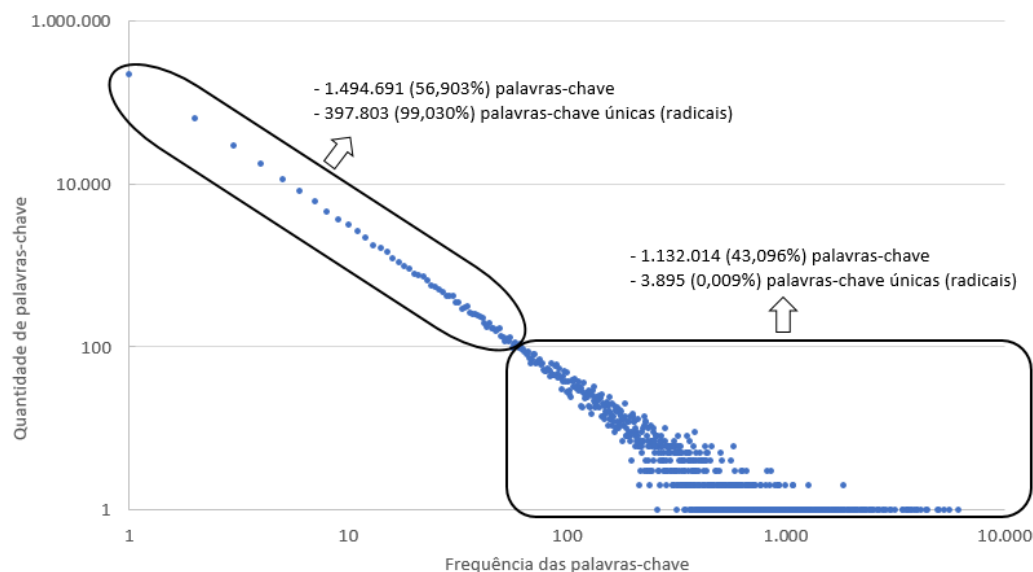


Figura 4. Distribuição das palavras-chave por sua frequência de aparecimento.

Assim, observa-se que 2.626.707 palavras-chave (conjunto global), após serem processadas, se reduziram a 401.698 palavras-chave únicas, comprovando a importância das etapas de tratamento dos dados para diminuição das informações a serem manipuladas. Também destaca-se que menos de 1% das palavras-chave únicas (equivalente a cerca de 43% do conjunto global) aparecem mais de 100 vezes, isto confirma a lei de Zipf, onde se descreve que um pequeno número de palavras é usado muito mais frequentemente.

Em continuação, para facilitar a compreensão da evolução temporal dos principais temas discutidos na grande área engenharias, todas as palavras-chave únicas foram ranqueadas de acordo com a medida de popularidade baseada na frequência a cada período de 5 anos entre 1962 e 2016. Assim, a Tabela 2 apresenta as principais palavras-chave de acordo com sua frequência de utilização em cada quinquênio entre todo o histórico analisado.

Tabela 2. As palavras-chave mais frequentes por quinquênios entre 1962 e 2016.

Ranque	1962-1966	Frequência	1967-1971	Frequência
1	Caulim	8	Ionosfera	12
2	Argilas Brasileiras	7	Conformação por Elastômero	8
3	Matérias-primas	7	Áreas Mineralizadas	7
4	Experimentos em Reator de Pesquisa	6	Corrosão	6
5	Corrosão sob Tensão	5	Sinterização	6
6	Argilas de Minas Gerais	5	Otimização	6
7	Caulinita	5	Planejamento Habitacional	6
8	Corrosão por Corrente de Fuga	4	Processos não convencionais	6
9	Corrosão em Frestas	4	Escoamento	5
10	Corrosão Intergranular	4	Recursos Hídricos	5
11	Análise Térmica	4	Urânio	5
12	Argilas de São Paulo	4	Experimentos em Reator de Pesquisa	5
13	Argilas Plásticas	4	Caulim	5
14	Haloisita	4	Projetos Sociais	5
15	Mineralogia	4	Agrogeologia	4
Ranque	1972-1976	Frequência	1977-1981	Frequência
1	Turbulências	20	Energia Solar	65
2	Recristalização	18	Otimização	55
3	Corrosão	14	Planejamento de Transportes	35
4	Dispositivo de Alta Pressão	14	Corrosão	33
5	Circuitos de Microondas	13	Energia	31
6	Energia Solar	13	Circuitos de Microondas	30
7	Conformação por Elastômero	12	Estruturas	28
8	Meios Porosos	12	Solidificação	28
9	Processos não convencionais	12	Meios Porosos	27
10	Caulim	11	Simulação	27
11	Alta Pressão	10	Aço Inoxidável	26
12	Dosimetria	10	Elementos Finitos	25
13	Otimização	10	Modelos Matemáticos	25
14	Atrito Interno	10	Propriedades Mecânicas	25
15	Carvão	9	Transferência de Calor	25
Ranque	1982-1986	Frequência	1987-1991	Frequência
1	Corrosão	99	Simulação	188
2	Elementos Finitos	81	Elementos Finitos	180
3	Simulação	81	Corrosão	158
4	Otimização	76	Microestrutura	156
5	Transferência de Calor	71	Cerâmica	133
6	Energia Solar	67	Métodos Elementos Finitos	121
7	Secagem	56	Transferência de Calor	120
8	Fadiga	52	Otimização	118
9	Leito Fluidizado	52	Propriedades Mecânicas	116
10	Modelos Matemáticos	51	Soldagem	107
11	Controle	47	Concreto Armado	106
12	Aço	46	Sinterização	100
13	Nióbio	46	Instrumentação	94
14	Instrumentação	45	Tratamento Térmico	88
15	Soldagem	44	Modelos	85

Ranque	1992-1996	Frequência	1997-2001	Frequência
1	Corrosão	519	Simulação	1.039
2	Simulação	447	Compósitos	1.015
3	Elementos Finitos	380	Otimização	938
4	Otimização	354	Modelagem	871
5	Qualidade	344	Elementos Finitos	819
6	Microestrutura	333	Corrosão	805
7	Compósitos	332	Ergonomia	794
8	Propriedades Mecânicas	330	Reciclagem	791
9	Modelagem	328	Concreto	783
10	Sinterização	276	Cerâmica	704
11	Cerâmica	261	Caracterização	668
12	Métodos Elementos Finitos	250	Qualidade	657
13	Alumínio	249	Redes Neurais	656
14	Concreto	249	Propriedades Mecânicas	651
15	Secagem	248	Microestrutura	634
Ranque	2002-2006	Frequência	2007-2011	Frequência
1	Compósitos	1.649	Biodiesel	2.169
2	Otimização	1.574	Compósitos	1.592
3	Reciclagem	1.504	Concreto	1.480
4	Concreto	1.471	Otimização	1.429
5	Simulação	1.349	Nanocompósitos	1.370
6	Modelagem	1.293	Meio Ambiente	1.199
7	Cerâmica	1.217	Biomateriais	1.148
8	Corrosão	1.215	Reciclagem	1.139
9	Resíduos Sólidos	1.132	Sustentável	1.139
10	Biomateriais	1.106	Corrosão	1.131
11	Elementos Finitos	1.063	Simulação	1.079
12	Caracterização	1.049	Resíduos Sólidos	1.005
13	Adsorção	1.045	Ergonomia	976
14	Propriedades Mecânicas	1.014	Qualidade da Água	954
15	Meio Ambiente	1.007	Resíduos	948
Ranque	2012-2016	Frequência		
1	Biodiesel	1.620		
2	Compósitos	1.423		
3	Sustentável	1.257		
4	Concreto	1.191		
5	Biomateriais	997		
6	Adsorção	994		
7	Otimização	980		
8	Corrosão	954		
9	Nanocompósitos	915		
10	Ergonomia	898		
11	Reciclagem	778		
12	Biomassa	777		
13	Simulação	776		
14	Resíduos	765		
15	Resíduos Sólidos	750		

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, exceto para os dois primeiros quinquênios analisados, onde apenas a palavra-chave “Caulim” define a similaridade dos temas discutidos para o período. Isso mostra que os interesses centrais dos doutores que publicam nas engenharias vão mudando gradualmente ao longo do tempo, e, pode estar relacionado a uma demanda da sociedade ou a maturação de um determinado tema.

Neste contexto, destacam-se a alta similaridade entre as principais palavras-chave estudadas nos quinquênios 2007-2011 e 2012-2016, onde os maiores interesses se diferem em

apenas duas palavras; sendo que, em 2007-2011, “Meio Ambiente” e “Qualidade da Água” estavam entre as principais, perdendo espaço para a “Adsorção” e “Biomassa” no quinquênio seguinte. Entretanto, quando comparados quinquênios sem ordem temporal, a similaridade entre os principais temas estudados diminui drasticamente, por exemplo, nenhum dos principais tópicos estudados em 1962-1966 foram evidenciados entre os primeiros a partir do quinquênio de 1977-1981, mostrando a troca completa de interesses dos temas estudados nas engenharias.

Os tópicos que tiveram mais presentes entre os principais durante todo o histórico analisado, foram: “Corrosão”, “Otimização” e “Simulação”. Esses três tópicos foram destaques em todos os quinquênios posteriores à primeira vez que apareceram em evidência, no caso, “Corrosão” teve em evidência pela primeira vez em 1967-1971, “Otimização” em 1972-1976 e “Simulação” em 1977-1981. Porém, a palavra-chave que teve maior frequência considerando todo o histórico foi “Compósitos”, aparecendo pela primeira vez entre as principais em 1992-1996 e, a partir do quinquênio seguinte, alternando entre primeiro e segundo lugares quanto a palavra-chave mais utilizada.

5. 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 em cada época. Ao analisar as palavras-chave dos currículos cadastrados na Plataforma Lattes é possível considerar publicações realizadas em anais de congressos, o que não é factível quando verificado repositório de dados internacionais ou em periódicos não indexados. Logo, é possível obter uma visão mais precisa dos tópicos mais investigados ao longo do tempo pelos pesquisadores brasileiros nas áreas de engenharias e, a partir disto, realizar análises acerca de fenômenos e tendências para extrair conhecimento a ser utilizado nos mais variados tipos de tomada de decisão no domínio científico.

Com base na medida de frequência, nota-se que um pequeno número de tópicos de pesquisas recebe maior atenção por parte dos doutores que publicam nas engenharias, enquanto uma grande quantidade de assuntos foram poucos estudados. Assim, pode-se perceber que as preferências centrais foram alterando gradualmente ao longo do tempo, possivelmente por uma demanda da sociedade ou maturidade de um determinado tópico de estudo.

Entretanto, apesar da validade dos resultados, percebe-se que a análise das palavras-chave utilizando apenas medida de importância de popularidade 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 equivocados. Com isso, como trabalhos futuros, espera-se que sejam realizadas análises que consideram fatores temporais para determinar a relevância de um tópico em cada época, análises baseadas em técnicas de redes sociais para determinar as palavras-chave mais centrais, e ainda, características qualitativas apropriadas ao contexto da Plataforma Lattes.

AGRADECIMENTOS

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

6. REFERÊNCIAS BIBLIOGRÁFICAS

- Brito, A.G.C., Quoniam, L. & Mena-Chalco, J. P., 2016. Exploração da Plataforma Lattes por assunto: proposta de metodologia. *Transinformação*, vol.28, n.1, pp. 77-86. Disponível em: <http://dx.doi.org/10.1590/2318-08892016002800006>
- Borges, V. A.; Nogueira, B.M.; Barbos, 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. Disponível em: <http://dx.doi.org/10.5753/rbie.2015.23.01.85>
- Choi, J., Yi, S. & Lee, K.C., 2011. Analysis of keyword networks in MIS research and implications for predicting knowledge evolution. *Information & Management*, vol. 48, n.8, pp. 371-381. Disponível em: <http://www.sciencedirect.com/science/article/pii/S0378720611000784>
- Dias, T. M. R., 2016. *Um estudo da produção científica brasileira a partir de dados da Plataforma Lattes*. Tese de Doutorado, Pós-Graduação em Modelagem Matemática e Computacional, Centro Federal de Educação Tecnológica de Minas Gerais (CEFET-MG).
- 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. ISSN 1808-5245. Disponível em: <http://seer.ufrgs.br/index.php/EmQuestao/article/view/53259>
- Dias, T. M. R.; Moita, G. F.; Dias, P. M; Moreira, T. H., 2014. Identificação e caracterização de redes científicas de dados curriculares. *Revista Brasileira de Sistemas de Informação*, vol.7, n.3, pp. 5-18.
- Digiampietri, L.A., 2015. *Análise da rede social brasileira*. Tese Livre Docência, Escola de Artes, Ciências e Humanidade, Universidade de São Paulo (USP).
- Ferraz, R.R.N., Quoniam, L. & Maccari, E.A., 2014. The use of ScriptLattes tool for extraction and on-line availability of academic production from a departament of stricto sensu in management. In *11th International Conference on Information Systems and Technology Management (CONTECSI)*, pp. 663-679.
- Ferreira, A. G. C, 2010. Bibliometria na avaliação de periódicos científicos. *Data Grama Zero. Revista de Ciência da Informação*, v. 11, n. 3, p. 1-9.
- Hong, Y.; Yao, Q.; Yang, Y.; Feng, J.J.; Wu, S.; Ji, W.; Yao, L.; Liu, Z., 2016. Knowledge structure and theme trends analysis on general practitioner research: A Co-word perspective. *Bio Med Central*, v. 17. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4734860>
- Kauer, A.; Moreira, V., 2013. Evolução dos Temas de Interesse do SBBD ao Longo dos Anos. *SBBD - Simpósio Brasileiro de Banco de Dados*, Recife, Brasil.
- Khan, G.F. & Wood, J., 2015. Information technology management domain: emerging themes and keyword analysis. *Scientometrics*, vol. 105, n. 2, pp. 959-972. Disponível em: <http://link.springer.com/article/10.1007/s11192-015-1712-5>
- Lucas, E. O.; Lara, M. L. G., 2014. Palavras-chave como elo entre artigos e autores: visualizações possíveis. In: *IV Encontro Brasileiro de Bibliometria e Cientometria*, Recife.
- Mccloskey, D. N, 1998. *The rhetoric of economics*. University of Wisconsin Press.

Mena-Chalco, J. P., Digiampietri, L.A., Lopes, F.M. & Junior, R.M.C., 2014. Brazilian bibliometric coauthorship networks. *Journal of the Association for Information Science and Technology*, vol. 65, n. 7, pp. 1424-1445. Disponível em: <http://dx.doi.org/10.1002/asi.23010>

Pritchard, A., 1969. *Statistical bibliography or bibliometrics?* *Journal of Documentation*, v.25, n. 4, pp. 348-349.

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

Souza, J. Júnior, M.L.M., Brito, A.V. & Duarte, A.N., 2014. Análise de redes de palavras baseada em títulos extraídos de um sistema de atendimento. In *III Brazilian Workshop on Social Network Analysis and Mining (BraSNAM)*, pp. 87-96.

Souza, J.; Júnior, M.L.M.; Brito, A.V. & DUARTE, A.N, 2015. 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)*. CSBC. Recife, Brasil. Disponível em: <http://dx.doi.org/10.13140/RG.2.1.3246.3207>

Vinkers, C. H.; Tjldink, J. K.; Otte, W. M., J., 2015. Use of positive and negative words in scientific PubMed abstracts between 1974 and 2014: retrospective analysis. *BMJ*. Disponível em: <http://www.bmj.com/content/351/bmj.h6467>

Yi, S. & Choi, J., 2012. The organization of scientific knowledge: the structural characteristics of keyword networks. *Scientometrics*, vol. 90, n. 3, pp. 1016-1026. Disponível em: <http://dx.doi.org/10.1007/s11192-011-0560-1>