

Identificação e Análise dos Principais Tópicos de Pesquisas em Engenharias no Brasil

Jether Gomes¹, Thiago M. R. Dias¹, Gray F. Moita¹

¹ CEFET-MG, jetherog@gmail.com

¹ CEFET-MG, thiago@div.cefetmg.br

¹ CEFET-MG, gray@dppg.cefetmg.br

¹ CEFET-MG - Centro Federal de Educação Tecnológica de Minas Gerais, Av. Amazonas, 7675, Nova Gameleira, 30510-000 Belo Horizonte, MG, Brasil.

Resumo. A ciência apresentou um crescimento impressionante nas últimas décadas quanto ao número de publicações de artigos científicos. Conseqüentemente, nota-se um esforço global de todas as áreas do conhecimento, negócios e fundações interessados em entender o patamar científico para, por exemplo, servir de base na construção de políticas científicas ou impulsionar grupos de pesquisas. Neste contexto, o objetivo deste trabalho é identificar e analisar temporalmente os principais tópicos de pesquisa publicados durante a trajetória da ciência brasileira por pesquisadores doutores que possuem currículos cadastrados na Plataforma Lattes e tem as Engenharias como grande área de atuação. 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 analisadas por métodos bibliométricos. Dentre estes dados, estão todas as palavras-chave de artigos publicados em periódicos entre 1997 e 2016, que são ranqueadas através da aplicação de uma medida de popularidade de tópico (TF.FI) proposta, que considera tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados.

Palavras-chave: Tópicos de Pesquisa, Palavra-chave, Bibliometria e 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 [9], a grande facilidade para acessar os artigos disponibilizados na internet contribui diretamente para a ampliação do conhecimento. Por outro lado, segundo Yi e Choi [8], a compreensão sobre a evolução da produção científica pode promover novos avanços na ciência. Em Brito et. al. [1], 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 ([9]; [6]; [3]; [5]; [10]; [11]). Neste ponto, é importante ressaltar que os trabalhos podem ser realizados sob diferentes perspectivas, como: 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 documento.

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 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. [7] 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 ([8]; [3]). Com isso, no atual trabalho, palavras-chave de artigos científicos são igualmente referenciadas como tópicos de pesquisas.

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

2 Desenvolvimento

A motivação da escolha da Plataforma Lattes como fonte de informação está relacionada basicamente a quatro fatores: (1) aos dados estarem disponíveis na internet e não terem sido amplamente analisados ([6]); (2) tratar da integração de dados de produções científicas de todas as áreas de C&T existentes na ciência brasileira ao longo de toda trajetória; (3) por não negligenciar os artigos publicados em periódicos nacionais que muitas vezes não são indexados, e também os artigos de anais de congresso ([9]); e, (4) por ser uma poderosa fonte para fornecimento de dados de alta qualidade para medir e avaliar o desempenho acadêmico nacional.

A aquisição dos currículos dos doutores que possuem Engenharias como área de atuação na versão XML foi realizada através da utilização do *LattesDataXplorer*, desenvolvido por Dias [9] para coletar os dados científicos contidos nos currículos cadastrados na Plataforma Lattes. Posteriormente, foi desenvolvido um arcabouço de componentes responsável por processar os dados dos currículos no intuito de suportar as análises desejadas (Figura 1). Nele, os componentes “filtragem dos dados” e “tratamento dos dados” são responsáveis por todo o processo de seleção, tratamento e modelagem das informações dos currículos que realmente necessitam ser processadas para atingir os objetivos propostos, e, concomitantemente, diminuir o tempo de processamento computacional.



Figura 1. 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 *ISCooll* proposto por Dias e Moita [11] para identificar colaborações científicas em grande volume de dados com custo linear. Assim, enquanto o método *ISCooll* original utiliza um dicionário para vincular os artigos (chaves de um dicionário) a seus autores (identificadores dos currículos), o método *ISCooll* 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).

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

Assim sendo, os dados referentes aos artigos dos doutores brasileiros contidos no arquivo de publicações científicas foram complementados com o valor do Fator de Impacto (FI) do respectivo periódico. Em seguida, o componente de “tratamento dos dados” foi ajustado para adicionar a somatória dos valores de FI dos periódicos por ano às respectivas palavras-chave dos artigos contidas nos arquivos de resultado padrão. Dessa maneira, as palavras-chave de artigos publicados em periódicos de relevância tem valor maior de FI em relação as palavras-chave de artigos publicados em periódicos de menor impacto. Isso foi possível através da utilização do conjunto de dados do JCR (Journal Citation Reports), pois, de posse do ISSN do periódico, permite obter informações referentes ao impacto e influência deste periódico na comunidade científica ano a ano desde 1997. A Figura 2 ilustra a adaptação necessária no arcabouço de componentes para a inserção do valor de FI dos periódicos no conjunto dos dados contidos no arquivo de publicações científicas e nos arquivos de resultado padrão.

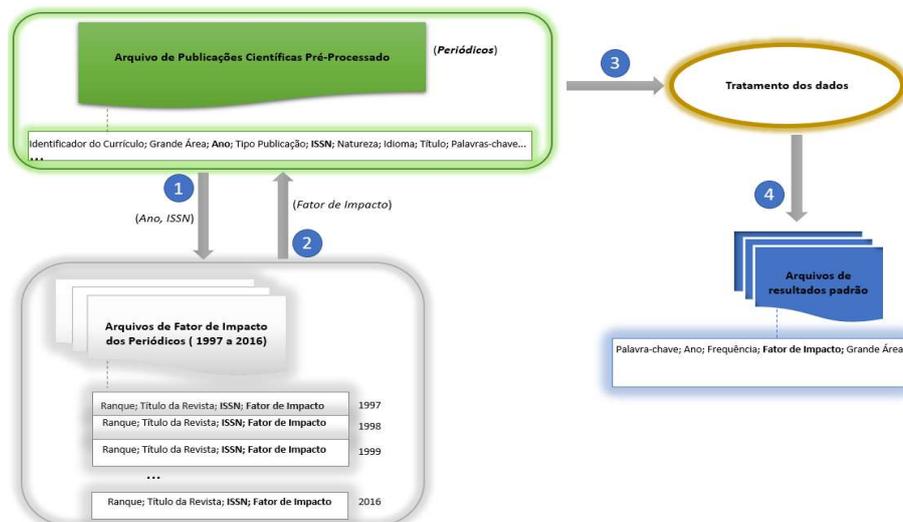


Figura 2. Processo para adição de FI dos periódicos ao conjunto dos dados científicos dos doutores.

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

Para possibilitar levar em consideração tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados, foi desenvolvida a medida de importância de tópico TF.FI. Nesse caso, a TF.FI pondera a frequência das palavras-chave de um conjunto de artigos publicados em periódicos com a soma dos valores de FI dos periódicos em que foram publicados os artigos que contém as respectivas palavras-chave. Essa medida parte do pressuposto que palavras-chave de um artigo publicado em um periódico de alta relevância são mais importantes que as mesmas palavras-chave de um outro artigo publicado em periódico de menor relevância, e, por isso, as palavras-chave de artigos publicados em periódicos de destaque recebem maior valor de FI como parcela de ponderação. A Figura 3 apresenta a equação para o cálculo da medida TF.FI de um determinado tópico num intervalo de tempo, para posteriormente esses tópicos serem ranqueados no intuito de destacar os mais importantes do conjunto de dados em análise.

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

Onde:

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

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

i = Ano inicial do período de análise

m = Ano final do período de análise

Figura 3. Medida de importância de tópicos TF.FI.

Para melhor entendimento, a Figura 4 apresenta um exemplo de utilização da medida de TF.FI a partir de um conjunto de dados de publicações científicas.

Autor	G. Área	Título da Publicação	Ano	Palavras-chave	Periódico	FI
...
ID 1	Engenharias	Título 1	2008	Cientometria; Redes Sociais; Tópicos de Pesquisa	0000-0001	10,00
ID 2	Engenharias	Título 2	2009	Bibliometria; Tópicos de Pesquisa	0000-0002	15,00
ID 3	Engenharias	Título 3	2012	Bibliometria; Redes Sociais; Tópicos de Pesquisa	0000-0002	20,00
ID 4	Engenharias	Título 4	2013	Bibliometria; Redes Sociais; Mineração de Texto	0000-0002	2,00
...

Palavras-chave	TF	FI	TF.FI
Tópicos de Pesquisa	3	45	135
Bibliometria	3	37	111
Redes Sociais	3	32	96
Cientometria	1	10	10
Mineração de Texto	1	2	2

Figura 4. Exemplo de aplicação da medida de importância TF.FI em um conjunto de dados científicos.

Como pode ser observado, “Tópicos de Pesquisa” tem popularidade maior que “Bibliometria” e “Redes Sociais”. Embora essas palavras-chave possuam o mesmo valor de índice de frequência, “Tópicos de Pesquisa” refere-se a artigos de periódicos de maior fator de impacto. Ou seja, para um determinado tópico de pesquisa possuir destaque a partir da aplicação da medida de importância TF.FI em um conjunto de dados, é necessário que este tópico tenha um considerável índice de frequência e que também esteja inserido em artigos publicados por periódicos com maior valor de fator de impacto em relação ao conjunto dos dados analisados.

3 Resultados e Discussões

Os dados foram coletados em abril de 2017, totalizando 265.170 currículos de doutores, em que 23.437 destes currículos têm as Engenharias como principal grande área de atuação. Para as análises, foram considerados todos os artigos únicos publicados pelos doutores em anais de congressos e em periódicos entre 1997 e 2016 referentes a grande área Engenharias, totalizando cerca 906.775 artigos e 2.626.707 palavras-chave.

Em continuação, para conhecer os tópicos que se constituem como os principais assuntos de pesquisa em engenharias no Brasil, todas as palavras-chave únicas foram ranqueadas de acordo com a medida de importância de popularidade TF.FI a cada período de 2 anos, conforme Tabela 1.

Tabela 1: As principais palavras-chave das Engenharias ranqueadas por TF.FI em cada biênio.

Ranque	1997-1998	TF.FI	1999-2000	TF.FI	2001-2002	TF.FI
1	Filme fino	560	Compósitos	3.434	Compósitos	4.928
2	Elementos Finitos	546	Modelagem	2.808	Filme fino	4.819
3	Modelagem	495	Filme fino	2.652	Cerâmicas	4.000
4	Otimização	450	Biomateriais	2.464	Microestrutura	3.876
5	Zeólito	406	Elementos Finitos	1.980	Modelagem	3.569
6	Alumina	352	Alumina	1.755	Cristalização	3.105
7	Transporte	348	Sol-Gel	1.450	Reologia	2.541
8	Compósitos	343	Cerâmicas	1.340	Biomateriais	2.340
9	Tokamak	342	Corrosão	1.140	Corrosão	2.320
10	Cerâmicas	308	Microestrutura	1.008	Varistor	2.128
11	Sno2	300	Cristalização	925	Permeabilidade	2.025
12	Corrosão	280	Redes Neurais	784	Processamento	1.920
13	Polianilina	266	Propriedade Mecânica	697	Elementos Finitos	1.764
14	Diamante CVD	265	Sinterização	660	Concreto	1.683
15	Varistor	264	Óculos	648	Redes Neurais	1.488
Total de Palavras-chave do Biênio		24.414		35.336		47.855
Ranque	2003-2004	TF.FI	2005-2006	TF.FI	2007-2008	TF.FI
1	Filme fino	13.110	Compósitos	17.751	Compósitos	18.998
2	Cerâmicas	12.441	Microestrutura	12.792	Cerâmicas	16.016
3	Compósitos	9.525	Cerâmicas	12.090	Filme fino	15.456
4	Microestrutura	8.910	Filme fino	6.834	Biomateriais	13.720
5	Biomateriais	5.529	Modelagem	6.264	Propriedade Mecânica	11.286
6	Modelagem	4.100	Propriedade Mecânica	6.188	Fotoluminescência	9.870
7	Nanopartícula	3.952	Otimização	5.967	Otimização	9.840
8	Otimização	3.434	Biomateriais	4.815	Biodiesel	9.118
9	Sinterização	3.174	Sensores	4.067	Nanocompósitos	8.633
10	Propriedade Mecânica	3.008	Cristalografia de Raios X	3.886	Cristalografia de Raios X	7.315
11	Lipase	2.744	Fotoluminescência	3.854	Adsorção	7.008
12	Fluído Magnético	2.610	Lipase	3.717	Microestrutura	6.723
13	Fotoluminescência	2.544	Corrosão	3.483	Modelagem	6.314
14	Cristalização	2.450	Adsorção	3.364	Lipase	6.097
15	Alumina	2.409	Redes Neurais	3.234	Corrosão	4.824
Total de Palavras-chave do Biênio		60.557		67.536		71.765
Ranque	2009-2010	TF.FI	2011-2012	TF.FI	2013-2014	TF.FI
1	Biodiesel	49.368	Biodiesel	130.634	Biodiesel	112.917
2	Nanocompósitos	28.032	Compósitos	28.310	Biomateriais	29.172
3	Compósitos	23.807	Nanocompósitos	25.758	Nanocompósitos	27.789
4	Biomateriais	20.150	Biomateriais	20.881	Biomassa	20.300
5	Otimização	12.998	Adsorção	16.107	Compósitos	20.230
6	Cerâmicas	12.480	Biomassa	15.000	Modelagem	19.685
7	Lipase	11.815	Lipase	12.782	Adsorção	17.696
8	Propriedade Mecânica	11.655	Propriedade Mecânica	12.096	Propriedade Mecânica	15.836
9	Modelagem	11.328	Modelagem	10.176	Microestrutura	12.760
10	Microestrutura	10.890	Quitosana	9.956	Nanopartícula	11.988
11	Adsorção	10.449	Otimização	9.628	Etanol	10.971
12	Fotoluminescência	7.888	Etanol	9.588	Corrosão	10.864
13	Biodegradável	7.705	Corrosão	9.559	Cerâmicas	10.692
14	Corrosão	6.532	Nanopartícula	8.772	Lipase	9.638
15	Alumina	6.384	Cerâmica	8.732	Otimização	7.600
Total de Palavras-chave do Biênio		80.820		86.105		90.713
Ranque	2015-2016	TF.FI	1997-2016	TF.FI		
1	Biodiesel	98.400	Biodiesel	1.891.960		
2	Compósitos	31.040	Compósitos	1.396.200		
3	Biomateriais	27.522	Biomateriais	1.057.614		
4	Biomassa	24.244	Cerâmicas	798.950		
5	Adsorção	23.868	Microestrutura	723.095		
6	Microestrutura	20.184	Nanocompósitos	693.666		
7	Nanopartícula	19.190	Modelagem	663.884		
8	Propriedade Mecânica	18.696	Propriedade Mecânica	634.848		
9	Nanocompósitos	18.592	Otimização	591.384		
10	Otimização	15.678	Adsorção	556.250		
11	Modelagem	10.890	Filme fino	492.600		
12	Modelo Matemático	9.559	Lipase	420.486		
13	Cerâmicas	9.545	Biomassa	407.037		
14	Etanol	9.159	Nanopartícula	353.363		
15	Simulação	7.663	Corrosão	347.422		
Total de Palavras-chave do Biênio		85.026		650.127		

Adicionalmente, a Tabela 2 apresenta o coeficiente de similaridade das 15 palavras-chave mais populares em cada par de biênios do período analisado. Neste contexto, coeficiente de similaridade mede o quão dois conjuntos de palavras-chave são iguais, independentemente da ordem do ranqueamento destas palavras-chave em cada conjunto.

Tabela 2: Coeficiente de similaridade das palavras-chave das Engenharias ranqueadas por TF.FI.

	1997-1998	1999-2000	2001-2002	2003-2004	2005-2006	2007-2008	2009-2010	2011-2012	2013-2014	2015-2016
1997-1998	X	46,66%	46,66%	40,00%	40,00%	40,00%	40,00%	26,66%	33,33%	26,66%
1999-2000		X	66,66%	66,66%	60,00%	53,33%	53,33%	33,33%	46,66%	40,00%
2001-2002			X	46,66%	53,33%	46,66%	40,00%	26,66%	40,00%	33,33%
2003-2004				X	66,66%	66,66%	66,66%	46,66%	60,00%	53,33%
2005-2006					X	86,66%	73,33%	53,33%	66,66%	53,33%
2007-2008						X	86,66%	66,66%	80,00%	66,66%
2009-2010							X	66,66%	80,00%	66,66%
2011-2012								X	86,66%	73,33%
2013-2014									X	86,66%
2015-2016										X

Por outro lado, a Tabela 3 apresenta os 15 principais tópicos estudados nas Engenharias considerando todo o período analisado.

Tabela 3: As principais palavras-chave das Engenharias ranqueadas por TF.FI entre 1997 e 2016.

Ranque	Palavras-chave (1997-2016)	TF.FI
1	Biodiesel	1.891.960
2	Compósitos	1.396.200
3	Biomateriais	1.057.614
4	Cerâmicas	798.950
5	Microestrutura	723.095
6	Nanocompósitos	693.666
7	Modelagem	663.884
8	Propriedade Mecânica	634.848
9	Otimização	591.384
10	Adsorção	556.250
11	Filme fino	492.600
12	Lipase	420.486
13	Biomassa	407.037
14	Nanopartícula	353.363
15	Corrosão	347.422

As Engenharias não apresentaram 100% de similaridade nos principais tópicos estudados entre os pares de biênios. Ou seja, a cada 2 anos os principais interesses dos doutores brasileiros que atuam nessa grande área mudaram. Contudo, ao analisar os pares de biênios sequentes, nota-se que em média 71% dos principais tópicos de pesquisas estudados no biênio atual foram relevantes no biênio anterior, destacando os pares de biênios 2005-2006 a 2007-2008; 2007-2008 a 2009-2010; 2011-2012 a 2013-2014 e 2013-2014 a 2015-2016, em que o percentual de similaridade foi de 86,66% (13 palavras-chave), havendo divergência, portanto, em apenas 2 dos 15 principais tópicos a cada 4 anos de estudo. Ao analisar os biênios não sequentes, é possível observar que 26,66% dos tópicos de pesquisas (4 palavras-chave) evidenciados como os principais no biênio de 2015-2016 também tiveram destaque no biênio 1997-1998 e, à medida que os principais tópicos de um determinado biênio são comparados com biênios mais recentes, o percentual de similaridade geralmente aumenta, mostrando que a alteração dos principais interesses dos doutores alternaram linearmente ao longo do tempo.

Dentre os principais tópicos elencados nas Engenharias entre 1997 e 2016, cerca de 86,66% (13 palavras-chave) dizem respeito a temas relacionados a engenharia dos materiais, sendo que 30% desses tópicos se referem a assuntos voltados para o desenvolvimento sustentável: “Biomassa”, “Biodiesel”, “Lipase” e “Biomateriais”. Nesse ponto, destaca-se “Biodiesel”, que apareceu pela primeira vez entre as principais palavras-chave no biênio 2007-2008 e atingiu o topo em popularidade em todos os biênios seguintes. Ainda é possível observar as palavras-chave “Modelagem” e “Otimização”, que possivelmente se relacionam à melhoria dos processos em um determinado domínio.

4 Conclusão

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 e atuam na grande área Engenharias. Ao realizar uma análise temporal entre as principais palavras-chave de cada período, foi possível verificar que as preferências centrais dos doutores sobre os tópicos de pesquisas foram alterando gradualmente ao longo do tempo, possivelmente devido a uma demanda externa da sociedade ou por determinado assunto ter atingido maturidade. Tais análises caracterizam-se como um importante mecanismo, pois podem identificar quais tópicos mais impactantes são desenvolvidos no Brasil e, assim, servir como apoio a diversos tipos de tomada de decisão. Como contribuição à validade dos resultados, é possível destacar a relevância do repositório da Plataforma Lattes para estudos bibliométricos sobre o desenvolvimento da ciência brasileira, aliado a abrangência e contemporaneidade de atualização dos dados considerados para análises.

A sistemática apresentada para a filtragem, tratamento e manipulação das palavras-chave contidas nos artigos de currículos da Plataforma Lattes forneceu uma metodologia nova para estudos sobre a identificação e a análises de tópicos de pesquisas. O mesmo procedimento pode ser replicado em outros conjuntos de palavras-chave de artigos científicos. Como alternativa às medidas de importância de tópicos com base na frequência e grau que consideram apenas atributos quantitativos, foi proposto uma medida de importância de tópico (TF.FI) que é capaz de levar em consideração tanto as características quantitativas das palavras-chave dos artigos quanto qualitativas dos periódicos em que tais artigos foram publicados. Assim, a medida de TF.FI consegue explorar em completude a heterogeneidade de níveis de impacto dos periódicos cadastrados na base curricular da Plataforma Lattes.

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

Autoria. Os autores confirmam que são os únicos responsáveis pela autoria deste trabalho, e que todo o material que foi incluído neste documento é propriedade (e autoria) dos autores, ou tem a permissão dos proprietários para para inclusão.

Referências

- [1] A. G. C. de Brito, L. Quoniam, and J. P. Mena-Chalco, “Exploração da plataforma lattes por assunto: proposta de metodologia,” *Transinformação-ISSN 2318-0889*, vol. 28, no. 1, 2016.
- [2] A. Pritchard, “Statistical bibliography or bibliometrics,” *Journal of documentation*, vol. 25, pp. 348, 1969.
- [3] G. F. Khan and J. Wood, “Information technology management domain: Emerging themes and keyword analysis,” *Scientometrics*, vol. 105, no. 2, pp. 959–972, 2015.
- [4] J. Choi, S. Yi, and K. C. Lee, “Analysis of keyword networks in mis research and implications for predicting knowledge evolution,” *Information & Management*, vol. 48, no. 8, pp. 371–381, 2011.
- [5] J. P. Mena-Chalco, L. A. Digiampietri, F. M. Lopes, and R. M. Cesar, “Brazilian bibliometric coauthorship networks,” *Journal of the Association for Information Science and Technology*, vol. 65, no. 7, 2014.
- [6] L. Digiampietri, “Análise da rede social brasileira,” Ph.D. dissertation, School of Arts, Sciences and Humanity, University of São Paulo (USP), 2015.
- [7] R. R. N. Ferraz, L. M. Quoniam, and E. A. Maccari, “The use of scriplattes tool for extraction and on-line availability of academic production from a department in management,” in *CONTECSI*, vol. 17, 2014.
- [8] S. Yi and J. Choi, “The organization of scientific knowledge: the structural characteristics of keyword networks,” *Scientometrics*, vol. 90, no. 3, pp. 1015–1026, 2012.
- [9] T. M. R. Dias, “Um estudo da produção científica brasileira a partir de dados da plataforma lattes,” Ph.D. dissertation, *CEFET-MG*, 2016.
- [10] T. M. R. Dias, G. F. Moita, P. M. Dias, and T. H. J. Moreira, “Caracterização de redes científicas de dados curriculares,” *iSys-Revista Brasileira de Sistemas de Informação*, vol. 7, no. 3, pp. 5–18, 2014.
- [11] T. M. R. Dias and G. F. Moita, P. M. Dias, “A method for the identification of collaboration in large scientific databases,” *Em Questão*, vol. 21, no. 2, pp. 140–161, 2015.