

Applying Text Mining to Identify Photovoltaic Technologies

A. de Lima, A. Argenta, I. Zattar, and M. Kleina

Abstract—Technologies for generating energy through fossil fuels have been increasingly being replaced. In this scenario, environmentally friendly innovations have gained space, one of the branches of renewable energy is solar energy that, through photovoltaic panels, converts the irradiation into electricity or heat. Therefore, the objective of this work is to present a patent analysis, carried out with the aid of text mining, through an algorithm developed for this purpose, using the R language, in order to identify the stage of technological development of photovoltaic panels. As results, we observed that there has been an increase in patent deposits in recent years, resulting in greater efficiency and lower cost of the modules. Innovations have been identified in several areas of knowledge, hybrid technologies (panels that generate electricity and heat), manufacturing methods, fast assembly systems and self-cleaning systems are among them.

Index Terms—Data mining, Language R, Patent, Photovoltaic technologies, Text mining.

I. INTRODUÇÃO

A rápida industrialização e urbanização aumentam acentuadamente o consumo de energia e produtos, resultando em altos índices de poluição [1]. Referência [2] ressaltam que nas últimas décadas houve um aumento no interesse de companhias e indústrias ao redor do mundo em desenvolver iniciativas sustentáveis. Dentro destas iniciativas, há a preocupação com a geração de energia oriunda de fontes renováveis como, por exemplo, a energia solar que fornece calor e eletricidade para inúmeras aplicações a um custo muito baixo. Contrastando com as fontes não renováveis, este tipo de fonte é ambientalmente amigável, pois produz menor emissão de poluentes. Portanto, é considerada uma das soluções mais sustentáveis para resolver a crise energética ao redor do mundo [3].

Embora já exista tecnologia desenvolvida para conversão da energia solar em térmica, o principal foco de pesquisa ainda recai sobre a sua conversão em eletricidade que, através de

módulos fotovoltaicos (FV's) capta uma maior quantidade de energia solar [3].

Em estudo de 2012, [4] ressaltam a oportunidade de utilização da tecnologia híbrida, composta por células solares fotovoltaicas e componentes solares térmicos. Esta combinação é capaz de, em um único módulo, gerar eletricidade e calor, melhorando a eficiência da conversão solar.

Para [5], dentre as tecnologias que auxiliam a maior eficiência dos módulos FV's estão os sistemas de rastreamento. Estes são capazes de aumentar a entrada de radiação solar, permitindo assim gerar mais energia. Contudo, por diferentes razões, sua implementação e instalação são mais complexas que os sistemas sem rastreadores [5]. Entretanto, conforme se pode observar nos documentos relativos a novas patentes, nos últimos anos houve um grande avanço neste tipo de tecnologia [6]. As patentes são capazes de mostrar a intensidade de movimentação tecnológica indicando as tendências do mercado. Segundo [7], a análise das patentes tem sido uma poderosa ferramenta analítica já empregada há muito tempo. Porém, devido ao elevado volume de documentos a serem analisados, pode-se tornar inviável a leitura individual. Para se contornar este problema, há uma infinidade de técnicas adotadas para a análise de patentes através da utilização de metadados. Estes dados fornecem informações como inventores, classificações, datas, citações, dentre outras [8].

Neste sentido, há uma crescente tentativa de integrar técnicas de mineração de dados, como a mineração de texto, *text mining*, para analisar oportunidades tecnológicas através das patentes [9]. A mineração de texto abrange várias ferramentas e técnicas, estas são utilizadas para identificar padrões e extrair conhecimento em linguagem não estruturada [10]. Para [11], a mineração de texto é capaz de mostrar as tendências a partir da análise de uma grande quantidade de dados textuais. Recentemente [8] desenvolveram um método com o objetivo de identificar padrões em patentes através da mineração de texto. Em suma, para os autores, dependendo do campo tecnológico a análise semântica se torna difícil devido às inúmeras formas de se escrever uma fórmula química por exemplo. Sua aplicação, aliada à consulta em bancos de dados que contêm informações tecnológicas, pode fornecer importantes indicadores, apresentando assim o estágio de desenvolvimento em determinado setor.

Neste contexto, este artigo tem por objetivo apresentar uma análise em patentes, realizada com o auxílio da mineração de texto, através de um algoritmo desenvolvido para este fim,

Este trabalho foi realizado com o apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Código de Financiamento 001.

A. I. de Lima, Universidade Federal do Paraná, Curitiba, PR, Brasil, andre.irazoqui@hotmail.com.

A. B. Argenta, Universidade Federal do Paraná, Curitiba, PR, Brasil, argenta_aline@yahoo.com.br.

I. C. Zattar, Universidade Federal do Paraná, Curitiba, PR, Brasil, izabel.zattar@gmail.com.

M. Kleina, Universidade Federal do Paraná, Curitiba, PR, Brasil, marianakleina11@gmail.com.

utilizando a linguagem R, a fim de identificar o estágio do desenvolvimento tecnológico de módulos FV's.

O artigo está organizado da seguinte maneira, a seção 2 apresenta os principais conceitos sobre a mineração de texto. A seção 3 descreve o protocolo de pesquisa e o desenvolvimento do algoritmo empregado. A seção 4 analisa os resultados obtidos. Finalmente, a última seção apresenta as conclusões, recomendações e sugestões para futuras pesquisas.

II. REVISÃO DA LITERATURA

A. Mineração de Texto

O volume de dados de fontes heterogêneas cresceu significativamente [12] e a produção científica está aumentando de forma consistente [13]. Identificar os dados relevantes a partir de uma quantidade enorme de informações é um desafio cada vez maior e, mais esforço é necessário para monitorar campos temáticos [12].

A mineração de texto surgiu com um propósito adequado para esta finalidade, pois oferece métodos para acessar e analisar fontes de dados textuais processando inúmeros documentos de forma automatizada [14], [15]. Neste sentido, [16] desenvolveram um sistema para extração de conhecimento em documentos não estruturados utilizando linguagem HTML.

Com a crescente produção textual na web, a área recebeu um forte impulso. Utiliza-se de técnicas para lidar com dados estruturados, semiestruturados ou não estruturados trabalhando com um banco de dados. Dentre as técnicas adotadas, incluem-se algoritmos para classificar, sumarizar, identificar entidades e *text clustering* [17].

Recentemente, [18] subdividiu a mineração de texto em seis etapas:

1. Definir o problema e objetivos;
2. Identificar e coletar o texto;
3. Organizar o texto;
4. Extrair as características;
5. Analisar;
6. Descrever as características da amostra.

Quanto ao tipo de abordagem, divide-se em duas categorias: Análise semântica e Análise estatística. A primeira emprega técnicas para avaliar a sequência dos termos no contexto do texto, fundamentando-se em técnicas de processamento de linguagem natural. Já a segunda, busca avaliar a importância do termo de acordo com a frequência que aparece no texto [19].

Para o cálculo da relevância, pode-se usar indicadores de Frequência dos termos (representa o número de vezes que um termo aparece em um documento), Frequência relativa (1) e Frequência inversa de documentos (2) além de outros.

$$F_{rel} = \frac{Frequência_{termo}}{Total_{palavras}} \quad (1)$$

$$IDF = \frac{Freq_i}{DocFreq_i} \quad (2)$$

Freq_i: número de vezes que o termo 't' aparece em determinado documento 'd'.

DocFreq_i: número de documentos que o termo 't' aparece.

Em relação a sua aplicação, a mineração de texto foi utilizada por [11] para identificar as tendências de sustentabilidade na indústria de processos químicos. Os autores analisaram 112 relatórios de sustentabilidade, do universo de 1.692.996 termos, que após aplicação dos filtros foram reduzidos para 219.516, foram identificadas as principais prioridades do setor.

Referência [12] salientam que a mineração de texto estende a base de conhecimento na análise de previsões. Os autores ressaltam que patentes e publicações científicas são comumente utilizadas como indicadores de mudança técnica. Estes documentos são encontrados em bancos de dados atualizados regularmente e, estas fontes são frequentemente utilizadas para prever e examinar o avanço tecnológico.

Para [20], avaliar e rastrear indicadores de sustentabilidade é desafiador. Estes estudos costumam consumir muito tempo e recursos financeiros. Com o objetivo de solucionar este problema, os autores buscaram explorar a viabilidade de identificar, rastrear e reportar os indicadores através da análise de artigos com dados não estruturados utilizando métodos de mineração de texto. No mesmo sentido [21] analisaram as mudanças na política nuclear da Coreia do Sul através destes métodos.

Assim, torna-se evidente a importância do emprego de ferramentas que auxiliem na análise de dados não estruturados.

III. METODOLOGIA

Nesta seção é apresentado o protocolo de pesquisa utilizado, o método de patentometria escolhido para o presente trabalho, e uma breve explicação sobre o algoritmo desenvolvido.

Tendo em vista o objetivo previsto neste artigo, que é realizar uma análise em patentes para identificar o desenvolvimento tecnológico de módulos FV's, o protocolo de pesquisa foi dividido em três macro fases. A primeira consistiu na definição do método (patentometria) e escolha da base patentária (Derwent Innovations Index® [22]). A patentometria, para alguns autores, pode ser considerada uma técnica oriunda dos métodos analíticos que pertencem à bibliometria. Os indicadores utilizados na análise de documentos de patentes sofrem adaptações de outros já utilizados em outros tipos de documentos [23]. Para [24], dentre outros estudos métricos, a patentometria é a que mais aproxima a academia com a indústria.

A segunda macro fase foi responsável pelo desenvolvimento do algoritmo, extração dos registros, limpeza dos dados e geração dos indicadores. Por último, na terceira macro fase, foi realizada a análise dos indicadores e conclusões.

Esta análise compreendeu o período de 2007 a 2016. Com o auxílio da mineração de texto foram consultados os títulos e resumos das patentes triádicas (patentes depositadas nos escritórios Europeu, Japonês e Norte Americano) da categoria correspondente ao código IPC (*International Patent Classification*) "H02S" - *GENERATION OF ELECTRIC POWER BY CONVERSION OF INFRA-RED RADIATION*,

VISIBLE LIGHT OR ULTRAVIOLET LIGHT, E.G. USING PHOTOVOLTAIC [PV] MODULES [25].

A. Desenvolvimento do Algoritmo em R

Algumas características presentes nos documentos de patentes são de suma importância neste tipo de pesquisa. Em primeiro lugar a padronização das informações nos documentos, seguido pela organização e classificação das patentes facilitando assim pesquisar sobre determinados nichos de mercado possibilitando estudar as publicações de países, empresas e instituições e, por último, mas não menos importante, a disponibilidade de dados em bases automatizadas [26]. Devido a esta padronização, parte dos dados pesquisados são estruturados e outra parte não estruturados. Portanto, a necessidade de desenvolver um algoritmo em R, advém da necessidade de interpretação e tratativa dos dados não estruturados.

Com um grande volume de documentos disponíveis para consulta, muitas vezes se torna inviável a leitura individual. Desta forma, foi desenvolvido um algoritmo em linguagem de programação R [27] utilizando os seguintes pacotes: tm [28], stringr [29] e dplyr [30], ggplot, ggplot2 [31] e wordcloud [32]. Estes pacotes são responsáveis por, respectivamente, minerar o texto, manipular as *strings* e gerar os gráficos.

A Fig. 1 ilustra a sequência de etapas realizadas no algoritmo.

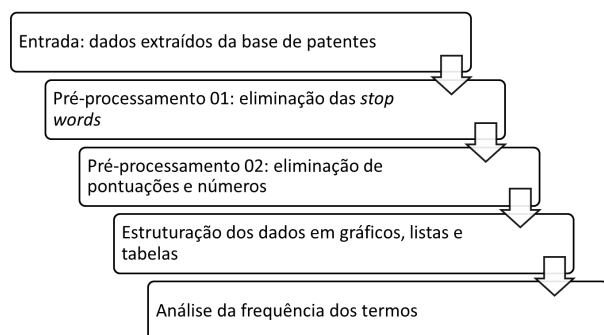


Fig. 1. Diagrama em blocos do algoritmo proposto.

Os dados exportados da base Derwent Innovations Index® foram salvos em um diretório. Para atender as etapas descritas na figura acima foram utilizados os códigos descritos nos procedimentos a seguir.

Procedimento 1 - Entrada

```
1. corpus<-VCorpus(DirSource("C:/Corpus",
encoding="UTF-8"))
, readerControl=list(reader=readPlain, language="eng") )
```

Procedimento 2 - Pré-processamento 01

```
1. corpus <- tm_map(corpus,
content_transformer(tolower))
2. corpus <- tm_map(corpus,removeWords,
stopwords("english"))
3. corpus <- tm_map(corpus,removeWords,
c("the", "can", "drawing", "THE",
"isidiidw", "usa","description", "woa",
```

```
"diidw","for","patent","aua","inventor",
"use"))
4. corpus <- tm_map(corpus,
stripWhitespace)
```

Procedimento 3 - Pré-processamento 02

```
1. corpus <- tm_map(corpus,
removePunctuation)
2. corpus <- tm_map(corpus, removeNumbers)
```

Procedimento 4 - Estruturação dos dados em gráficos listas e tabelas

```
1. freq <- TermDocumentMatrix(corpus)
2. matriz <- as.matrix(freq)
3. matriz <- sort(rowSums(matriz),decreasing=T)
4. matriz <- data.frame
(word=names(matriz),freq=matriz)
5. barplot(matriz$freq[20:1],
names.arg=matriz$word[20:1], horiz=T,
axis.lty=10, las=2, cex.names = 0.7)
6. wordcloud(corpus, max.words=100,
random.order=T, colors=rainbow(8), rot.per=
0.5,use.r.layout=T)
```

Procedimento 5 - Análise da frequência dos termos

```
1. dtm <- DocumentTermMatrix(corpus)
2. inspect(dtm)
3. findFreqTerms(dtm, 200)
4. findAssocs(dtm, "termo", 0.8)
5. inspect(removeSparseTerms(dtm, 0.4))
6. inspect(DocumentTermMatrix(corpus,
list(dictionary= c("termoA","termoB"))))
```

IV. RESULTADOS

Os resultados apresentados a seguir, buscam identificar o estado atual de desenvolvimento tecnológico de módulos FV's, através de uma análise de patentes, sendo oriundos tanto de dados estruturados (volume de depósitos, nome dos depositantes e áreas do conhecimento), os quais são obtidos diretamente da base patentária, como não estruturados (informações presentes no título e resumo), os quais foram obtidos através do uso de programação feita em R.

O volume de depósitos de patentes em determinada categoria indica a intensidade de movimentação tecnológica na área, sendo que ao longo de todo o período analisado foram encontrados 889 documentos depositados concomitantemente nos escritórios norte americano, japonês e europeu. No primeiro ano do período da análise (2007) até o ano de 2011 houve aumento discreto no número de depósitos, porém a partir de 2012 o crescimento se mostrou bastante agressivo, Fig. 2. Provavelmente, esse salto no número de depósitos é reflexo dos esforços que países vêm fazendo para fomentar o desenvolvimento das tecnologias ambientalmente amigáveis.

No ano de 2010, a OMPI (Organização Mundial da Propriedade Intelectual) lançou o chamado IPC Green Inventory, esta ferramenta online contém a relação das tecnologias consideradas ambientalmente amigáveis. O objetivo deste inventário é facilitar a busca por informações relacionadas a estas tecnologias. Desde então, alguns países

têm criado mecanismos para estimular o desenvolvimento de tecnologias sustentáveis, dentre eles podemos citar os programas de Patentes Verdes [33].

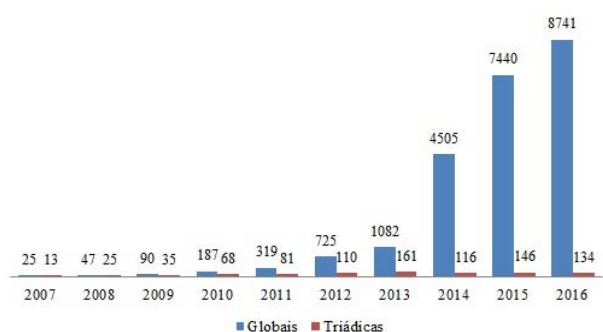


Fig. 2. Depósitos de patentes por ano.

Inaugurado antes mesmo do IPC Green Inventory, o primeiro programa de patentes verdes é de origem britânica, no ano de 2009. Seguido pelos programas da Austrália (2009), Coreia do Sul (2009), Japão (2009), Estados Unidos (2009), Israel (2009), Canadá (2011), Brasil (2012), China (2012) e Letônia (2013). Estes programas têm por objetivo acelerar o exame das patentes que neles se enquadrem [33], [34]. Além do mais, outras políticas de incentivo foram adotadas, tais como redução de impostos, políticas ambientais e linhas de financiamentos voltadas ao desenvolvimento sustentável [35].

De 2007 até 2011 o número acumulado de depósitos das patentes triádicas foi de 222, sendo equivalente a aproximadamente 33% do total. Entre os anos de 2012 e 2013 foram 271 depósitos no grupo (15%). Já de 2014 até 2016 o volume de depósitos neste grupo de escritórios atingiu 396 (2% do total). Este fenômeno se deve muito à participação chinesa no desenvolvimento da tecnologia. Quando adicionadas as patentes depositadas no escritório chinês, passamos a ter um total de 16341 depósitos neste último período, englobando 79% do total. A Fig. 3 detalha o número de depósitos de patentes triádicas acumuladas ao longo de todo o período pesquisado.

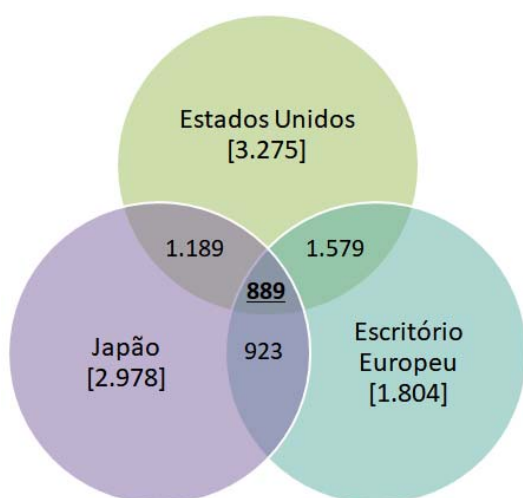


Fig. 3. Diagrama de Venn acumulado (2007 – 2016).

A análise por área de conhecimento mostra uma uniformidade ao longo de todo o período, com destaque para: engenharia, instrumentos de instrumentação e combustíveis energéticos. Vale salientar que, ao longo do período analisado, as áreas ciências poliméricas e ciência dos materiais estão em ascensão, Fig. 4.

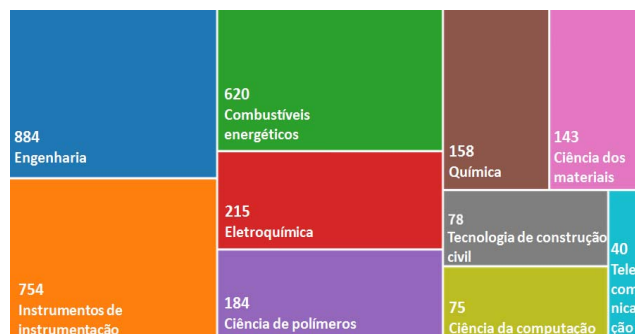


Fig. 4. Áreas de conhecimento (2007 – 2016).

Dentro da área de instrumentação o foco está em aumentar a eficiência dos módulos, através do equilíbrio entre a eficiência de conversão e eficiência de rastreamento (EP2993754-A1, JP2017527001-W, US2017264099-A1). Já as ciências poliméricas e ciência dos materiais têm buscado aumentar a resistência ao impacto e à variação de temperatura (JP2016186046-A, EP3275936-A1, US2018066134-A1).

Após o tratamento dos dados não estruturados no algoritmo, foram identificadas 67704 palavras. Os termos mais frequentes da amostra podem ser conferidos na Fig. 5. A matriz com a frequência absoluta dos termos possui 10 linhas (representando os anos da amostra) e 4807 colunas (termos presentes nos arquivos).

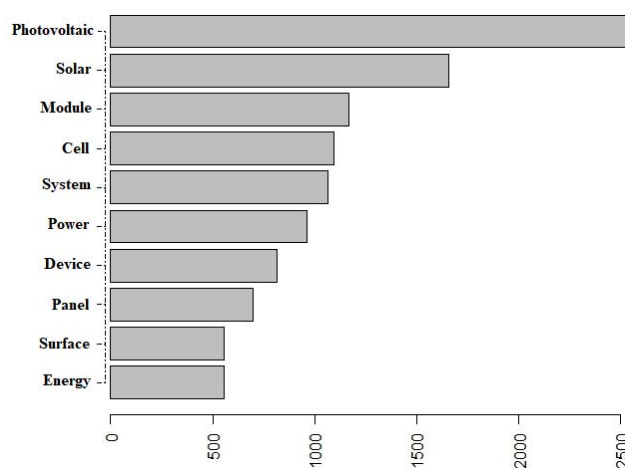


Fig. 5. Frequência dos termos da amostra (acumulado (2007 – 2016)).

A análise da frequência dos termos sofre poucas alterações a partir de 2010. Porém o crescimento de determinados termos se destaca entre os demais. Este fenômeno pode ser observado com as palavras *heat* e *tracking* a partir de 2011 (chegando a alcançar 0,44% e 1,15% de frequência relativa em determinados anos na análise), corroborando com a tese de [4] de que havia oportunidade para exploração da tecnologia

híbrida. Ademais, patentes relacionadas ao termo *heat* também trazem tecnologias para dissipação do calor, tornando os módulos mais eficientes, resistentes e confiáveis (US8563847-B2, JP5966027-B2, EP2443666-A4). A nuvem de palavras apresenta os 100 termos mais frequentes dentro de todo o período analisado, Fig. 6.

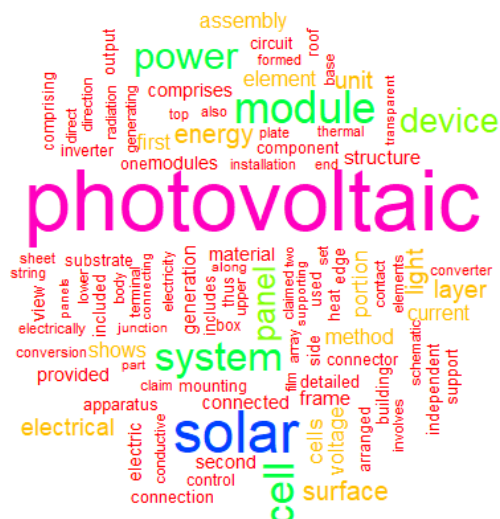


Fig. 6. Nuvem de palavras.

Em relação aos códigos IPC que possuem maior relevância na amostra, o subgrupo H01L-031/042 (Módulos fotovoltaicos ou matrizes de células fotovoltaicas individuais) aparece com maior frequência, Fig. 7. Na sequência aparece o código H01L-031/048 (Encapsulamento de módulos). A partir de 2012 a subclasse H02S (geração de energia elétrica por conversão de radiação infravermelha, luz visível ou luz ultravioleta, e.g. usando módulos fotovoltaicos [PV] (obtido energia elétrica de fontes radioativas G21H 1/12; dispositivos semicondutores inorgânicos sensíveis à luz H01L 31/00; dispositivos termoeletrônicos H01L 35/00; dispositivos piroeletrônicos H01L 37/00; dispositivos semicondutores orgânicos sensíveis à luz H01L 51 / 42)) passa a ter forte presença.

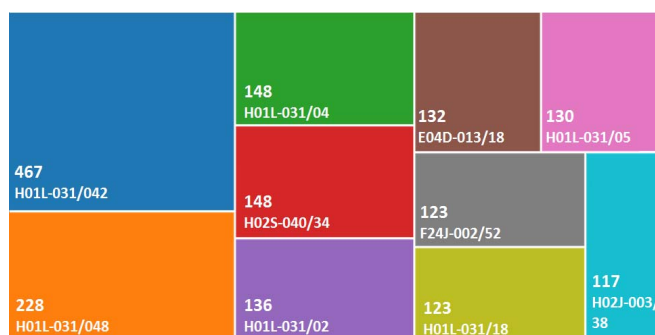


Fig. 7. Depósitos por código IPC (2007 – 2016).

Dentre as empresas que se destacam (Fig. 8), a norte americana Sunpower Corp é a que possui o maior número de depósitos. A empresa tem desenvolvido inúmeras tecnologias, desde sistemas autolimpantes (US2015349706-A1,

JP2017514430-W, EP3151984-A4) que otimizam a utilização do líquido de limpeza reduzindo custos de manutenção possibilitando um menor número de operadores, sistemas de acoplamento que permitem montagem rápida para redução do custo de instalação (tanto para uso residencial quanto comercial) (US2016261227-A1, JP2018508669-W, EP3266102-A4) até sistemas de rastreamento térmico para módulos FV's (US8796535-B2, EP2761676-A1, JP2014534797-W).



Fig. 8. Empresas depositantes (2007 – 2016).

Outro destaque entre os depositantes é a Dow Global Technologies, também de origem norte americana. As tecnologias depositadas englobam tanto dispositivos quanto métodos de fabricação (US8163125-B2, EP2464516-B1, JP5647247-B2). Uma das preocupações da empresa é resolver o problema de disparidade nas propriedades de material das camadas em painéis multicamadas, sendo uma das causas à contração/expansão térmica.

A análise da frequência dos termos também mostra o crescimento do termo *concentrator*, sua presença está relacionada a patentes dos CPV's (*Concentrator Photovoltaic Panels*). O CPV é um ramo da tecnologia fotovoltaica, onde são utilizados concentradores ópticos de luz para coletar e concentrar a irradiação solar incidente em uma grande área em pequenas células solares [36].

V. CONCLUSÃO

As tecnologias de módulos fotovoltaicos têm demonstrado forte intensidade no volume de depósitos de patentes nos últimos anos. A eficiência dos módulos está aumentando devido a inúmeros aspectos, as patentes mostram que os esforços de empresas e pesquisadores estão voltados ao desenvolvimento de novos dispositivos de rastreamento, novas técnicas de fabricação, novos materiais, novos sistemas de montagem e sistemas autolimpantes. Como resultado, além da maior eficiência, os módulos tendem a ser cada vez mais economicamente viáveis, pois seus custos de montagem e manutenção se tornarão menores conforme estas inovações cheguem ao mercado.

A aplicação da mineração de texto auxiliou na identificação dos termos mais frequentes nos documentos. O algoritmo em linguagem de programação R se mostrou muito preciso na análise. A análise do conteúdo, através dos indicadores gerados com a mineração de texto, possibilitou identificar as tendências presentes nos documentos.

Como sugestão para trabalhos futuros, analisar o desenvolvimento de tecnologias específicas dos módulos FV's.

REFERÊNCIAS

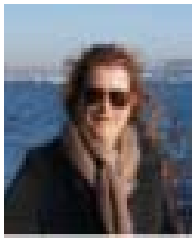
- [1] Z. Cheng, "The spatial correlation and interaction between manufacturing agglomeration and environmental pollution," *Ecol. Indic.*, vol. 61, pp. 1024–1032, 2016.
- [2] F. Ali, C. Boks, and N. Bey, "Design for Sustainability and Project Management Literature - A Review," *Procedia CIRP*, vol. 48, pp. 28–33, 2016.
- [3] A. Syafiq, A. K. Pandey, N. N. Adzman, and N. A. Rahim, "Advances in approaches and methods for self-cleaning of solar photovoltaic panels," *Sol. Energy*, vol. 162, no. May 2017, pp. 597–619, 2018.
- [4] X. Zhang, X. Zhao, S. Smith, J. Xu, and X. Yu, "Review of R&D progress and practical application of the solar photovoltaic/thermal (PV/T) technologies," *Renew. Sustain. Energy Rev.*, vol. 16, no. 1, pp. 599–617, 2012.
- [5] N. AL-Rousan, N. A. M. Isa, and M. K. M. Desa, "Advances in solar photovoltaic tracking systems: A review," *Renew. Sustain. Energy Rev.*, vol. 82, no. August 2017, pp. 2548–2569, 2018.
- [6] P. Gonçalves *et al.*, "Photovoltaic technologies: Mapping from patent analysis," *Renew. Sustain. Energy Rev.*, vol. 93, no. April, pp. 215–224, 2018.
- [7] K. Song, K. S. Kim, and S. Lee, "Discovering new technology opportunities based on patents: Text-mining and F-term analysis," *Technovation*, vol. 60–61, no. January, pp. 1–14, 2017.
- [8] H. Niemann, M. G. Mochrle, and J. Frischkorn, "Use of a new patent text-mining and visualization method for identifying patenting patterns over time: Concept, method and test application," *Technol. Forecast. Soc. Change*, vol. 115, pp. 210–220, 2017.
- [9] C. Lee, B. Kang, and J. Shin, "Novelty-focused patent mapping for technology opportunity analysis," *Technol. Forecast. Soc. Change*, vol. 90, no. PB, pp. 355–365, 2015.
- [10] R. Paynter *et al.*, "EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews EPC Methods: An Exploration of the Use of Text-Mining Software in Systematic Reviews," vol. 84, pp. 33–36, 2017.
- [11] W. Te Liew, A. Adhitya, and R. Srinivasan, "Sustainability trends in the process industries: A text mining-based analysis," *Comput. Ind.*, vol. 65, no. 3, pp. 393–400, 2014.
- [12] V. Kayser and K. Blind, "Extending the knowledge base of foresight: The contribution of text mining," *Technol. Forecast. Soc. Change*, vol. 116, pp. 208–215, 2017.
- [13] L. Bornmann and R. Mutz, "Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references," *J. Assoc. Inf. Sci. Technol. Growth*, no. 284, p. 8, 2014.
- [14] S. M. Weiss, N. Indurkha, T. Zhang, and F. Damerou, *Text Mining Predictive Methods for Analyzing Unstructured Information*. Springer-Verlag New York, 2005.
- [15] R. Feldman and J. Sanger, *The text mining handbook*. CAMBRIDGE UNIVERSITY PRESS, 2007.
- [16] T. Rodriguez and J. Aguilar, "Knowledge Extraction System from Unstructured Documents Knowledge Extraction System from Unstructured Documents," *IEEE Lat. Am. Trans.*, vol. 16, no. 2, pp. 639–646, 2018.
- [17] P. Oliveira Lima Junior, L. Gonzaga de Castro Junior, and A. Luiz Zambalde, "Applying Textmining to Classify News About Supply and Demand in the Coffee Market," *IEEE Lat. Am. Trans.*, vol. 14, no. 12, pp. 4768–4774, 2016.
- [18] T. Kwartler, *Text Mining in Practice with R*. 2017.
- [19] E. A. M. Morais and A. P. L. Ambrósio, "Mineração de Textos," 2007.
- [20] S. J. Rivera, B. S. Minsker, D. B. Work, and D. Roth, "A text mining framework for advancing sustainability indicators," *Environ. Model. Softw.*, vol. 62, pp. 128–138, 2014.
- [21] C. Park and T. Yong, "Prospect of Korean nuclear policy change through text mining," *Energy Procedia*, vol. 128, pp. 72–78, 2017.
- [22] Clarivate Analytics, "Derwent Innovations Index," 2018. [Online]. Available: http://apps-woofknowledge.ez22.periodicos.capes.gov.br/DIIDW_GeneralSearch_input.do?product=DIIDW&search_mode=GeneralSearch&SID=5DdGRoa47Pyhj8phMCw&preferencesSaved=.
- [23] M. C. Fagundes, P. A. de A. Garcia, G. D. S. Motta, and D. R. Armond-de-Melo, "Perfil Tecnológico Da Csn: Um Estudo Patentométrico," *Rev. Adm. Innov. - RAI*, vol. 11, no. 1, p. 276, 2014.
- [24] C. A. Pereira, "O fluxo e as dimensões socioespacial e socioinstitucional do conhecimento em Ciência, Tecnologia & Inovação: um estudo patentométrico da produção tecnológica da Unicamp," Pontifícia Universidade Católica de Campinas, 2008.
- [25] WIPO, "IPC Green Inventory," 2019. [Online]. Available: https://www.wipo.int/classifications/ipc/en/green_inventory/.
- [26] M. V. Gúzman-Sánchez, "Patentometria: herramienta para el análisis de oportunidades tecnológicas," Universidad de La Habana, 1999.
- [27] R. C. Team, "R: A language and environment for statistical computing." R Foundation for Statistical Computing, Vienna, Austria, 2018.
- [28] I. Feinerer and K. Hornik, "tm: Text Mining Package. R package version 0.7-5." 2018.
- [29] H. Wickham, "stringr: Simple, Consistent Wrappers for Common String Operations." 2018.
- [30] H. Wickham, R. François, L. Henry, and K. Müller, "dplyr: A Grammar of Data Manipulation." 2018.
- [31] H. Wickham, "ggplot2: Elegant Graphics for Data Analysis." 2016.
- [32] I. Fellows, "wordcloud: Word Clouds." 2018.
- [33] A. Dechezleprêtre, "Fast-Tracking 'Green' Patent Applications: An Empirical Analysis," *Cent. Econ. Perform.*, no. 1197, p. 27, 2013.
- [34] P. C. Reis, C. C. Osawa, M. E. M. Martinez, J. C. C. B. R. Moreira, and D. A. Santos, "Programa das Patentes Verdes no Brasil: Aliança Verde entre o Desenvolvimento Tecnológico, Crescimento Econômico e a Degradação Ambiental," pp. 1–17, 2016.
- [35] S. Zeng, Y. Liu, C. Liu, and X. Nan, "A review of renewable energy investment in the BRICS countries: History, models, problems and solutions," *Renew. Sustain. Energy Rev.*, vol. 74, no. March 2016, pp. 860–872, 2017.
- [36] M. C. Browne, B. Norton, and S. J. McCormack, "Heat retention of a photovoltaic/thermal collector with PCM," *Sol. Energy*, vol. 133, pp. 533–548, 2016.



André Irazoqui de Lima possui graduação em Engenharia de Produção pela Universidade Federal do Pampa, Bagé/RS, Brasil em 2013, MBA em Gestão Empresarial pela Fundação Getúlio Vargas, Rio de Janeiro/RJ, Brasil em 2016 e mestrado em Engenharia de Produção pelo Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal do Paraná, Curitiba/PR, Brasil em 2019. Atualmente atua como consultor de empresas pela AIL Engenharia & Consultoria, Bagé, Rio Grande do Sul, Brasil.



Aline Brum Argenta possui graduação em Engenharia de Alimentos pela Universidade Federal do Pampa, Bagé/RS, Brasil em 2013, mestrado em Engenharia de Processos pela Universidade Federal de Santa Maria, Santa Maria/RS, Brasil em 2015 e atualmente é doutoranda em Engenharia de Alimentos no Programa de Pós-Graduação em Engenharia de Alimentos na Universidade Federal do Paraná, Curitiba/PR, Brasil.



Izabel Cristina Zattar possui graduação em Tecnologia em Mecânica pela Sociedade Educacional de Santa Catarina, Santa Catarina, Brasil em 2002, mestrado em Engenharia Mecânica pela Universidade Federal de Santa Catarina em 2004 e doutorado em Engenharia Mecânica pela Universidade Federal de Santa Catarina, Florianópolis/SC, Brasil

em 2008. Atualmente desenvolve trabalhos voltados à Inovação como professora e pesquisadora na Universidade Federal do Paraná, com parcerias consolidadas com Institutos de Tecnologia, Incubadoras de Base Tecnológica e Agências de Fomento.



Mariana Kleina possui graduação em Matemática Industrial, Curitiba/PR, Brasil em 2009, mestrado em Métodos Numéricos em Engenharia pela Universidade Federal do Paraná, Curitiba/PR, Brasil em 2012 e doutorado em Métodos Numéricos em Engenharia pela Universidade Federal do Paraná, Curitiba/PR, Brasil em 2015. Tem

experiência nas áreas de Pesquisa Operacional e Inteligência Artificial. Atualmente é professora e pesquisadora na Universidade Federal do Paraná.