

folha de rosto

Revista de Biblioteconomia e Ciência da Informação

Fontes de Informação em Patentes: análise das características das bases Derwent Innovations Index, ORBIT, INPI, Google Patents e PatentScope com base na produção tecnológica da UFRGS

Ana Maria Mielniczuk de Moura

Doutora em Informação e Comunicação, pela Universidade Federal do Rio Grande do Sul (UFRGS). Docente do Programa de Pós-Graduação em Ciência da Informação (PPGCIN) da UFRGS.

ana.mmoura@uol.com.br

Fernanda Bochi dos Santos

Mestre em Ciência da Informação, pela Universidade Federal do Rio Grande do Sul (UFRGS). Doutoranda em Ciência da Informação, pelo Programa de Pós-Graduação em Ciência da Informação da Universidade Estadual Paulista (UNESP).

nanda.bochi@gmail.com

Ana Paula Medeiros Magnus

Mestranda em Informação e Comunicação, pela Universidade Federal do Rio Grande do Sul (UFRGS).

magnus@gmail.com

Letícia Angheben El Ammar Consoni

Mestre em Informação e Comunicação, pela Universidade Federal do Rio Grande do Sul (UFRGS). Bibliotecária da Escola de Engenharia da UFRGS.

le.angheben@gmail.com

Resumo

A pesquisa explora as diferentes bases de dados depositantes de patentes — importantes fontes de informação para a realização de estudos patentométricos — e propõe-se a investigar as características inerentes ao processo de coleta e recuperação de patentes depositadas nas bases de dados Derwent Innovations Index, Orbit, INPI, Google Patents e PatentScope. Compreende a necessidade de discriminar as diferenças nessas bases de dados a partir de um único objeto de estudo, reunido na produção intelectual da Universidade Federal do Rio Grande do Sul (UFRGS). De caráter qualitativo-descritivo, o estudo tem por objetivo explorar as bases de dados mencionadas. A busca e a recuperação dos dados das patentes depositadas pela UFRGS e indexadas nas bases ocorreu pela utilização variada do nome da universidade, visando a abranger o maior número de registros possível. Observou-se que as bases de dados analisadas apresentaram características distintas umas das outras. A DII e a Orbit exibiram o maior número de registros recuperados, porém a DII recuperou registros indexados de forma incorreta. O INPI foi a base com o número de registros mais fidedigno à Universidade, da mesma forma que a PatentScope. O Google Patents ficou muito aquém das expectativas, não servindo como uma boa fonte de dados para o tipo de pesquisa realizado, pois não indexa o escritório brasileiro. Conclui-se que é necessário examinar atentamente a fonte de informação antes de realizar um estudo de cunho patentométrico, pois existem divergências e características singulares entre as bases de dados.

Palavras-chave: Cientometria. Patentometria. Fontes de informação. Base de dados. Patentes.

Evaluation of the Librarianship Course of the Federal University of Ceará in the perception of the graduates

Abstract

The research explores the different patent depository databases — important sources of information for conducting patentometric studies — and proposes to investigate the inherent characteristics of the process of collection and retrieval of patents deposited in the databases DII, Orbit, INPI, Google Patents and PatentScope. It comprehends the need to discriminate the differences in these databases from a single object of study, gathered in the intellectual production of the Universidade Federal do Rio Grande do Sul (UFRGS). Of qualitative-descriptive character, the study aims to explore the databases mentioned. The search and recovery of the data of the patents deposited by the UFRGS and indexed in the bases occurred by the varied use of the name of the university, aiming to cover the largest possible number of records. It was observed that the databases analyzed presented different characteristics from each other. DII and Orbit exhibited the highest number of retrieved records, but DII retrieved indexed records incorrectly. The INPI was the basis with the most reliable number of records to the university, in the same way as PatentScope. Google Patents was away from expectations, not serving as a good source of data for the type of research done, as it does not index the Brazilian office. It is concluded that it is

René Faustino Gabriel Júnior

Doutor em Ciência da Informação, pela Universidade Estadual Paulista Júlio de Mesquita Filho (Unesp).

renefgj@gmail.com

necessary to closely examine the source of the information before carrying out a patentometric study, since there are divergences and singular characteristics among the databases.

Keywords: Scientometrics. Patentometry. Information sources. Database. Patents.



Este trabalho está licenciado com uma Licença [Creative Commons - Atribuição-NãoComercial-Compartilhalqual 3.0 Brasil](https://creativecommons.org/licenses/by-nc-sa/3.0/br/).

1 Introdução

Uma das principais características da sociedade atual é a capacidade de produzir informações em uma velocidade imensurável. Os avanços científicos e tecnológicos estão sendo desenvolvidos mais rápido que a capacidade das pessoas de processarem tais informações. De acordo com Caldas e Silva (2016), a produção de dados científicos nada mais era do que dados com medidas entre *bit* e *byte*. Hoje se fala em *exabytes* de dados, que, por sua vez, são gerados diariamente, chegando a acumular, no período de um ano, quase dois *zettabytes* de dados — números que crescem a cada instante com o avanço tecnológico, provocando, com isso, uma necessidade de compreensão dos instrumentos que promovem a busca, a coleta e a análise dos dados pesquisados. A publicação de artigos científicos e o patenteamento das novas descobertas acabam convergindo com esse panorama de aumento do fluxo informacional, o que torna necessária a compreensão de como esse aumento vem caracterizando-se no âmbito científico-tecnológico.

Como forma de organizar o crescente aumento da informação, surgiram as bases de dados, caracterizadas como a expressão utilizada para indicar a coleção de dados que serve de suporte a um sistema de recuperação de informações (CUNHA, 2001). Souto (2003) define bases de dados como sistemas elaborados, que têm o objetivo de fornecer informação atualizada, precisa e confiável, buscando atender à demanda de uma clientela específica, diferenciando-se por suas diversas características, como: abrangência, proporcionalidade geográfica, materialidade da informação, formato dos documentos, tipo de conteúdo (especializado ou geral), tipologia (bases de referência, resumos ou de documentos na íntegra) e acesso (livre ou via assinatura).

As bases de dados que indexam documentos de patentes contribuem para a elaboração de uma patente, pois é imprescindível realizar a pesquisa de anterioridade de uma tecnologia para formalizar um pedido de patente. Além disso, essas bases contêm dados que apoiam a pesquisa científica, mostrando o que está sendo inventado, bem como trazendo informações que podem servir de suporte para identificar as áreas tecnológicas mais fomentadas, as instituições ou empresas que detêm um número considerável de patentes, os campos tecnológicos mais estudados e os nichos de mercados mais explorados (JAGHER, 2010; GHESTI, 2016).

Para uma melhor percepção do processo de recuperação da informação dos dados indexados em bases de dados, faz-se necessário o entendimento das características de recuperabilidade dessas informações. A disciplina que estuda a recuperação da informação ocupa-se de conceitos como a descrição da informação e suas especificidades, visando à busca e à acessibilidade de dados pertinentes a uma pesquisa (MONTEIRO; FERNANDES, DECARLI, TREVISAN et al., 2017). No contexto da Ciência da Informação, a recuperação da informação aponta para a necessidade de utilização de termos e metadados padronizados, para que o processo de descoberta da base seja o mais preciso possível. A padronização das

informações bibliográficas disponíveis nas bases de dados, muitas vezes, não segue uma tendência global, sendo formadas por uma série de fatores que vão desde políticas de cada desenvolvedor/empresa em termos de gestão dos dados até a funcionalidade social (SOUZA, 2006).

Assim como os artigos científicos, o documento de patente segue uma tendência global e distinta de descrição, é formado por diversos elementos bibliográficos padronizados, como nome dos inventores, depositantes, área do conhecimento, a partir do uso da Classificação Internacional de Patentes (CIP) e vinculação a uma família de patentes, assim como a descrição da invenção requerida e sua aplicação industrial (ARAÚJO, 1981; FRANÇA, 2000). O texto que forma o documento é redigido de maneira técnica e pontual, com o objetivo de reproduzir a invenção e, também, facilitar a transcrição do mesmo para outros idiomas, caso seja de interesse patentear a invenção em outros países (GARCIA, 2004). Compreende-se, ainda, que os documentos de patentes podem ser categorizados com função dupla, apresentando valoração além de seu caráter de informação especializada, mas, também, como símbolo de inovação e desenvolvimento tecnológico e econômico de um país (GARCIA, 2004; MOURA; ROZADOS; CAREGNATO, 2006; MUELLER; PERUCCHI, 2014; CONSONI, 2017).

Decorrente do que foi exposto, esta pesquisa tem como problematização saber: quais as características inerentes ao processo de coleta e recuperação de patentes indexadas nas bases de dados Derwent Innovations Index (DII), Orbit Intelligence (Orbit), Instituto Nacional de Propriedade Industrial (INPI), Google Patents e PatentScope? Compreendendo o impacto do crescimento e da velocidade de produção da informação e sabendo que, como fonte de pesquisas de patentes, existem várias bases de dados disponíveis, porém ainda se desconhece suas coberturas, suas formas de recuperação, indexação e padronização dos dados, justifica-se o desenvolvimento deste estudo. O presente trabalho visa a dar luz aos estudos patentométricos, apresentando as diferenças e semelhanças nas bases de dados utilizadas para alcançar os objetivos propostos, procurando identificar as características dos recursos e dos mecanismos de busca, aferir sobre a cobertura, indexação, recuperação e padronização das bases estudadas.

Visando a propor um critério único para investigar as bases de dados expostas, optou-se por verificar a produção tecnológica em forma de patentes da Universidade Federal do Rio Grande do Sul (UFRGS), indexadas nas bases de dados. A escolha dessa instituição como objeto do estudo baseia-se pela Universidade ter sido ranqueada como uma das dez melhores universidades federais brasileiras, segundo os dados do último Índice Geral de Cursos (IGC) de 2016 (atualizado em abril de 2018), que avalia as instituições de ensino superior no país e é realizado pelo Ministério da Educação (MEC) (BRASIL, 2018). Trata-se de uma universidade que vem crescendo ao longo dos anos e investindo em ciência e tecnologia, reiterando o que já fora mencionado pela escolha do objeto.

2 Metodologia

A presente pesquisa trata de um estudo descritivo que visa a comparar as características de cobertura, indexação, recuperação e padronização das bases de dados utilizadas para a realização de estudos patentométricos. As bases selecionadas para a realização desta pesquisa contemplam bases de dados que são comumente utilizadas em estudos dessa natureza, como DII, Orbit, Google Patents, INPI e PatentScope.

A coleta dos dados foi realizada ao longo do ano de 2018, e foi estabelecido um critério único para a efetivação do estudo, isto é, optou-se por utilizar um mesmo objeto de análise em todas as bases, uma vez que o intuito da pesquisa é verificar as diferenças e semelhanças entre elas. Para isso, decidiu-se pesquisar pela produção tecnológica da UFRGS. As expressões de busca foram testadas em cada uma das bases de dados com as variações do nome da Universidade, sem limitação temporal predeterminada, possibilitando buscar todas as patentes que foram indexadas. Porém, tendo em vista que nem todas as variações possibilitam recuperação, as expressões que apresentaram dados pertinentes à pesquisa são detalhadas nos resultados. Por fim, os dados coletados foram registrados em uma planilha Excel, para possibilitar as análises, e as observações das características de cada base foram identificadas durante o processo de utilização das mesmas.

3 Resultados

A pesquisa foi realizada nas bases de dados DII, Orbit, INPI, Google Patents e PatentScope, com diferentes estratégias de busca. Foi recuperado um total de 1.767 patentes nas quatro bases analisadas, conforme a **Tabela 1**.

Tabela 1 – Resultados obtidos nas bases de dados pesquisadas e expressões de busca utilizadas

Base de Dados	Resultados	Expressões de Buscas
ORBIT	428	(Universidade Federal do Rio Grande do Sul)
DII	410	AN=(UNIV AND FEDERAL AND RIO AND GRANDE AND SUL) OR AN=(UNIV AND FEDERAL AND DO AND RIA AND GRANDE AND DO AND SUL AND UFRGS) OR AN=(UNIV AND FEDERAL AND DO AND RIO AND GRANDE AND DO AND SUL) OR AN=(Hospital AND Clinicas AND PA) OR AN=(Hospital AND Clinicas AND Porto AND Alegre) OR AN=(Agron AND Univ AND Fed AND Rio AND Grande AND do AND Sul) OR AN=(Cenbiot AND UFRGS) OR AN=(Fdn AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fdn AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Hosp AND Rio AND Grande AND do AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND do AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul AND State) OR AN=(Fed AND Univ AND Rio AND Grande AND Sul AND UFRGS) OR AN=(Fed AND Univ AND RS) OR AN=(Federal AND University AND Rio AND Grande AND do AND Sul) OR AN=(HCPA) OR AN=(Rio AND Grande AND do AND Sul AND Fed AND Univ) OR AN=(UFRGS) OR AN=(UFRGS) OR AN=(Univ AND Fed AND Rio AND Do AND Sul) OR AN=(Univ AND Fed AND Rio AND Grande AND do AND Sul) OR AN=(Univ AND Fed AND Rio AND Grande AND Sul) OR AN=(URFGS)
Google Patents	33	"UFRGS" OR "Universidade Federal Do Rio Grande Do Sul -UFRGS" OR "Universidade Federal Do Ria Grande Do Sul -UFRGS" OR "Federal University of Rio Grande Do Sul" OR "Univ Federal Do Rio Grande Do Sul - Ufrgs"
INPI	404	UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
PatentScope	492	Universidade Federal do Rio Grande do Sul
TOTAL	1.767	

Fonte: Dados da Pesquisa.

Em relação à cobertura das bases de dados, a única que não indexa patentes registradas no escritório brasileiro (INPI) é a Google Patents, pois não faz parte dos 17 escritórios mundiais cobertos pela sua busca. Contudo, os resultados obtidos nessa base refletem os países/escritórios internacionais em que a UFRGS também registrou seus pedidos de patentes — por isso, torna-se importante, também, o uso dessa base de dados. De acordo com o **Quadro 1**, observa-se as diferenças de resultados obtidos e cobertura nas quatro bases de dados.

Quadro 1 – Patentes da UFRGS recuperadas nas bases de dados DII, Orbit, INPI, Google Patents e PatentScope

	DII	Orbit	INPI	Google Patents	PatentScope
Registros recuperados	410	428	404	33	492
Registros válidos	390	321	404	33	366
Patentes com sigla BR	373	318	404	0	269
Patentes concedidas	390	241	326	33	97
Patente mais recente nos registros	07/06/2017	21/11/2017	09/01/2018	01/08/2017	17/07/2018
Patente mais antiga nos registros	26/08/1997	26/06/1990	04/01/1990	02/09/2005	26/04/2000

Fonte: Dados da Pesquisa.

Os registros que aparecem em “registros válidos”, destacados no **Quadro 1**, são aqueles que resultaram após a exclusão dos registros duplicados e a conferência da titularidade da patente (ou seja, somente foram considerados os registros nos quais o nome da Universidade constava como depositante da patente). Essa ocorrência foi observada nos resultados da DII, Orbit e PatentScope.

Alguns erros foram identificados nos registros recuperados na DII — por exemplo, a atribuição da Patente BR102016008747-A2 como sendo a depositária a UFRGS, sendo que, no INPI, consta como depositária a Universidade Federal do Rio Grande do Norte (UFRN). Além disso, a padronização do nome das instituições na DII não é controlada. Dentre as variantes do nome da UFRGS, foram localizadas, também, as seguintes variações: “UNIV FEDERAL DO RIO GRANDE DO SUL UFRGS”, “UFRGS INST FISICA”, “UNIV FEDERAL DO RIA GRANDE DO SUL UFRGS (UYRI-Non-standard)” (destaque nosso), “UNIV FEDERAL DO RIO GRANDE DO SUL (UYRI-Non-standard)”, “UNIV FEDERAL RIO GRANDE DO SUL (UYRI-Non-standard)”, “UNIV FEDERAL RIO GRANDE DO SUL (UYRI-Non-standard)”, “UNIV FEDERAL RIO GRANDE DO SUL UFRGS (UYFE-Non-standard)” (destaque nosso). Destaca-se, ainda, que alguns registros estão com erros de grafia e outros incorporam a sigla da instituição. Na base de dados Orbit, também foram encontradas variações do nome da UFRGS, estando registrada como “UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL” e “UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL - UFRGS” (destaque nosso).

Em relação às características dos recursos e dos mecanismos de buscas identificados nas bases de dados, a DII é uma base de patentes que indexa documentos provenientes de mais de 40 escritórios nacionais, internacionais e regionais. Além do acesso a mais de 71 milhões de patentes, cobertas desde 1963 (CLARIVATE ANALYTICS, 2017), a indexação feita pela DII acrescenta outras informações ao registro, que facilitam a busca e a recuperação de documentos, como o Derwent Class Code (que busca e recupera uma determinada categoria de invenção) e o Derwent Manual Code (que indica novos aspectos técnicos da invenção). As buscas podem ser realizadas por meio de vários rótulos, ou seja, várias possibilidades de campos de pesquisa. Os registros recuperados também incluem links para os documentos citados (que podem ser outras patentes, artigos de periódicos ou outro documento), atribuídos pelos examinadores das patentes. Todas essas características auxiliam na análise dos dados quantitativos de citação, bem como na busca e recuperação das informações relacionadas à inovação. Ao longo dos anos, a DII tem sido utilizada em diversos estudos patentométricos que envolvem a análise dos dados estatísticos de patentes (OKUBO, 1997; SPINAK, 2003; MARICATO; NORONHA, 2013; CONSONI; MOURA, 2016; MAGNUS, 2018). Isso se deve, principalmente, à característica abrangente do seu conteúdo e à existência de mecanismos de exportação dos dados para a posterior análise. Para esta pesquisa, o acesso ao conteúdo ocorreu por meio do Portal de Periódicos Capes.

A Orbit é uma base de dados paga, produzida pela Questel Orbit desde a década de 1970, e disponibilizada por meio de assinatura, porém não está acessível pelo Portal da CAPES. Oferece um sistema de busca, seleção, análise e exportação de informações contidas em mais de 55 milhões de famílias de patentes e mais de 81 milhões de patentes aplicadas em 100 países e escritórios de registros de patentes. Ela apresenta dados sobre depositantes, inventores, textos com traduções

para diversos idiomas, imagens e citações, reunidos em bases de dados proprietárias: FamPat e FullPat (respectivamente, família de patentes e pedidos de patentes). As principais funcionalidades da base podem ser assim agrupadas: busca, visualização, seleção, análise, exportação, monitoramento e compartilhamento. Contém uma tabela de família de patentes que permite personalizar e exibir valores métricos para cada família em seu conjunto de análise. É possível escolher as métricas mais relevantes para uma pesquisa, classificar por valores de métricas para identificar as patentes-chave e exportar o conjunto de dados completo em vários formatos de arquivos, viabilizando, assim, a realização de análises detalhadas.

O INPI disponibiliza, em seu site, um recurso de busca, em que é possível pesquisar as patentes por diferentes campos, podendo mesclá-los, mas não permitindo cruzar as informações ou aplicar estratégias de buscas mais elaboradas. Nesse recurso, não há a possibilidade de exportação dos dados, dificultando a recuperação de grandes quantidades de registros. Por isso, desde agosto de 2017, de forma a se adequar ao Decreto nº 8.777/2016, o INPI passou a disponibilizar, no Portal de Dados Abertos, a Revista de Propriedade Industrial (RPI), com a possibilidade de baixar os metadados das seções da revista, contendo informações sobre o registro de solicitações, os recursos e as concessões de patentes. A RPI se encontra disponível online e é publicada semanalmente às terças-feiras pelo INPI, em uma numeração sequencial e incremental. A revista está estruturada em oito seções: I) comunicados; II) contratos de tecnologia; III) desenhos industriais; IV) indicações geográficas; V) marcas; VI) patentes; VII) programas de computador; e VIII) topografia de circuito integrado (BRASIL, 2017). Por ter características diferentes de apresentação dos dados, para fins desta pesquisa, foram baixados os arquivos da RPI de P1132.ZIP, edição de 11/08/1992 até o P2450.ZIP, edição de 19/12/2017. Esse recurso permite identificar o nome e endereço URL de todas as RPIs disponíveis em formato PDF, ZIP+TXT¹ e/ou ZIP+XML até 24/Jan/2017.

O Google Patents é uma base de acesso gratuito que permite a busca e leitura de textos completos de patentes de vários escritórios de depósitos de patentes. Disponibiliza um conjunto de mais de 90 milhões de patentes de 17 escritórios mundiais de registros de patentes, tornando esses dados públicos por meio do BigQuery, que é uma ferramenta de armazenamento de dados (WETHERBEE, 2017). A interface de busca é simplificada, similar à busca do Google. Além disso, a base disponibiliza a ferramenta *Search and read the full text of patents from around the world* (Pesquise e leia o texto completo das patentes de todo o mundo), que permite visualizar em tempo real a quantidade de patentes concedidas e aplicadas em cada um dos 17 escritórios cobertos. Oferece, também, o recurso de pesquisa em documentos no Google Scholar (*Include non-patent literature*), permitindo a busca de anterioridade e estado da arte daquilo que se deseja patentear. O Google Patents, por meio de classificação automática, utiliza a Classificação Cooperativa de Patentes, para que os resultados sejam recuperados (GOOGLE, c2018).

A PatentScope é uma base de dados vinculada ao WIPO (World Intellectual Property Organization), um dos principais escritórios de patentes do mundo, e tem acesso gratuito, permitindo a busca e a leitura dos documentos disponibilizados. A interface está disponível tanto no idioma inglês como em português, sua usabilidade é simples e de fácil acesso aos documentos. A base conta com ferramentas que auxiliam a tradução dos metadados para nove idiomas, incluindo o português. Há possibilidade de acesso às informações básicas dos documentos. Sua cobertura conta com mais de 73 milhões de documentos e outros 3,4 milhões de pedidos de patentes, abrangendo cerca de 56 entidades, mescladas entre países e escritórios de patentes. Os documentos de patentes brasileiras estão inseridos na base desde 1972 e, atualmente, somam um montante de aproximadamente 737.288 documentos (WORLD INTELLECTUAL PROPERTY ORGANIZATION, 2018).

O **Quadro 2** apresenta as principais características da recuperação e padronização dos dados nas diferentes bases de dados analisadas.

Quadro 2 – Características da recuperação e padronização dos dados das bases DII, Orbit, INPI, Google Patents e PatentScope

	DII	Orbit	INPI	Google Patents	PatentScope
Cobertura de tempo	1963 - presente	1960 - presente	1970 - presente	1970 - presente	1972 - presente
Língua	Interface em português. Descrições dos registros em inglês.	Interface em inglês.	Interface em português.	Interface em inglês.	Interface em inglês.
Apresentação dos resultados	Número da patente, título, inventores, depositantes e número de acesso primário na DII. Quando a patente está disponibilizada na íntegra, é possível baixar o documento, clicando no botão "original". Possibilita a visualização de até 50 registros por página.	Número de família de patente, número prioritário, título, inventores, depositantes, PDF do documento original, imagens, resultado dos dados, recursos de leitura e seleção de resultados, associadas a ferramentas de exportação de dados, com hyperlinks para textos integrais, citações, palavras-chave no contexto (KWIC), ferramenta de <i>highlight</i> (colorir termos de busca), gráficos de citações, gráficos de famílias de patentes, ferramenta de comparação de reivindicações, ferramenta de reordenamento dos resultados.	Número da patente, datas de solicitação, título da patente e sua classificação. Para acesso aos dados completos, é necessário acessar o número da patente. Visualização de 20 registros por página.	Número da CIP, título, dados do inventor e depositante, datas de depósito, publicação e indexação na base, pequena parte do resumo, possibilidade de realizar outra busca com o mesmo número de CIP da patente recuperada. Possibilita a visualização de até 100 registros por página.	Número de patente, título da patente, depositante, requerente, datas de depósito, CIP, resumo.
Forma de exportação dos dados	De 500 em 500 registros.	Permite exportar em diversos formatos, sem limite no número de exportação. Além disso, permite exportar: texto, imagem, campos selecionados, hyperlinks permanentes (imagens, textos integrais pdf, citações, dados da família de patentes).	No site de busca, não existe exportação de dados. A partir de agosto de 2017, os dados da RPI estão no portal Dados Abertos do Governo Federal, com todas suas seções.	Exporta todos os resultados em uma única vez.	Mediante cadastro, é possível exportar todos os registros recuperados.
Formato de exportação dos dados	html ou .txt, com possibilidade de conversão do arquivo .txt para Excel, tanto para Windows como para Macintosh.	.txt, .pdf, .xls, .xlsx, .csv, .xml	Não aplicado. Para o portal de dados abertos, os dados estão em .txt.	.cvs	.xls
Informações contidas na exportação	Número da patente (PN), título (TI), inventor (AU), depositante (AN), data de indexação (GA), resumo (AB), resumo expandido (TF), código de classe da DII (DC), código manual da DII (MC), CIP (IP), ano de publicação (PD), ano de depósito (AD), patentes citadas (CP), artigos citados (CR).	Título (TI), Resumo (AB), Classificação Internacional ou Classificação Europeia ou CPC (IC/EC/CPC), Basic Index (BI) Super-Abstract (SA), número de publicação (PN), Número de publicação de família (NPN), Número Prioritário de família (NPR), Data da primeira publicação da família de patentes (EPD).	(21) N° do Pedido; (22) Data do Depósito; (43) Data da Publicação; (47) Data da Concessão; (51) Classificação IPC; (54) Título; (57) Resumo; (71) Nome do Depositante; (72) Nome do Inventor; (74) Nome do Procurador.	Número da patente (ID), título (Title), depositante (Assignee), inventor (Inventor /author), data de depósito (Priority date), data de indexação (Filing/creation date), data de publicação (Publication date), data de concessão	ID da aplicação; Número de aplicação; Número de Publicação; Data de publicação; Título; Dados prioritários; IPC; Requerentes; Inventores; Imagem FP; Entradas Nat Phase.

				(Grant date) e o link para o registro na base.	
Campos de busca	Busca simples e avançada com os campos: assunto, título, inventor, depositante, número da patente, CIP, código de classe na DII, código manual na DII, número de acesso primário na DII, código do depositante, nome mais código do depositante, número de patente citada.	Apresenta diferentes formas de busca, tais como: busca básica, busca por menus, busca por linhas de comando avançadas, busca por citações, busca por similaridade, busca por famílias de patentes expandidas, assistente de busca semântica, assistente de busca por números de publicação, assistente de busca por titulares, assistentes de busca por classificações, buscas por conjuntos de resultados previamente selecionados.	Busca simples e avançada com os campos: número do pedido, número do recolhimento, do protocolo, data de depósito, data de prioridade, data de publicação, número de classificação, palavra-chave pelo número de classificação, palavra-chave no título e no resumo, nome do depositante, CPF ou CNPJ do depositante e nome do inventor.	Busca simples e avançada com os campos: termos, datas de depósito, inventor, depositante, escritório, língua, status e tipo.	Busca simples e avançada, permite combinação dos campos, expansão em outros idiomas, pesquisa por compostos químicos mediante login.
Disponibilidade de filtros na base	Sim	Sim	Não	Não	Sim
Dados de citação	Links para as citações dos examinadores.	Hyperlinks permanentes de citações.	Não se aplica.	Dados das patentes que foram citadas no documento e das patentes que já citaram a patente que está sendo visualizada.	Não se aplica.
Recursos de análise dos resultados	Recurso "Análise de Resultados" com contagem do número de registros por depositantes, inventores, CIPs e área do conhecimento.	Contagem do número de depositantes, inventores e CIPs do resultado da pesquisa. Apresenta gráficos, grafos, tabelas, análise da evolução temporal das subclasses internacionais, mapas geográficos.	Não se aplica.	Contagem do número de depositantes, inventores e CIPs do resultado da pesquisa. Mostra um gráfico com a contagem relativa dos 5 primeiros resultados em relação ao ano de indexação na base.	Na aba "Filtro", aparece um resumo dos registros recuperados por país, principal CIP, principal inventor, principal requerente e data da publicação.

Fonte: Dados da Pesquisa.

4 Considerações Finais

Os estudos de patentes, ou patentométricos, estão sendo cada vez mais comuns na academia, buscando a compreensão de como se apresentam as relações entre a ciência e a tecnologia. Muito embora as bases de dados não possam ser comparadas por apresentarem atributos únicos e serem produzidas por empresas diferentes, ainda assim é relevante o seu estudo, por serem consideradas ferramentas que disponibilizam dados que auxiliam pesquisas científicas. Nesse contexto, esta pesquisa teve como objetivo analisar as características de algumas fontes de informações que disponibilizam dados sobre patentes, possibilitando um melhor entendimento das particularidades de cada uma, sua abrangência e suas singularidades. Para esse fim, utilizou-se a UFRGS como elemento de análise nas bases para realização do levantamento das características que foram analisadas.

Observou-se que cada base de dados apresenta-se de forma distinta, seja por sua abrangência, cobertura, escritórios que indexa, formas de recuperação, de exportação e tempo de atualização. Analisando atentamente, pode-se considerar que as bases têm finalidades de consultas diferentes, se considerarmos as necessidades informacionais de quem realizar as buscas. Embora elas sejam desenvolvidas para atender um público específico – os desenvolvedores de novas tecnologias, também são bastante utilizadas por pesquisadores com interesses em estudos métricos voltados aos documentos de patentes e às secretarias e aos núcleos de desenvolvimento tecnológico das universidades.

Em relação à cobertura do INPI, observou-se que as patentes depositadas pela UFRGS tiveram a melhor cobertura, além de seguirem um padrão na descrição de metadados, como o nome dos requerentes. Este resultado é compreensível, pois reflete o fato de o órgão ser o regulador da propriedade industrial no país, sendo sua base uma das principais fontes de consulta para os residentes que desejam patentear suas invenções.

Já as bases DII e Orbit apresentam documentos de diversos países. Essa característica de apresentação de documentos depositados em escritórios nacionais, internacionais e regionais pode ser mais interessante se a finalidade do usuário for depositar sua invenção em outros países nos quais não seja residente. Observou-se que, na DII, há falta de padronização no nome de inventores e depositantes, não somente na grafia, mas também nos registros de forma trocada de nomes das instituições depositantes. Essa imprecisão dos metadados por parte da DII faz com que a busca na base seja mais demorada, já que exige do usuário um controle dos termos utilizados para a realização da pesquisa e uma conferência dos registros. Quanto à Orbit, esta se caracterizou como sendo a que apresenta melhor qualidade dos dados, com atualização periódica e disponibilização de análises detalhadas dos resultados.

A base PatentScope apresentou resultados semelhantes ao INPI, mesmo tendo um número de registros menor que a base brasileira. Nesse ponto, ressalta-se a semelhança entre essas duas bases, ambas fontes de informação que agem muito mais como fonte de consulta e registro da documentação. Ou seja, apresentam a finalidade de disponibilizar os dados e divulgar os documentos que estejam indexados dentro das situações legais.

Dentre as fontes pesquisadas, a base que apresentou os resultados mais insatisfatórios foi o Google Patents, no contexto das patentes depositadas pela UFRGS. Para verificar se o mesmo resultado aplica-se a outros objetos, tornam-se necessárias pesquisas mais aprofundadas na base. O principal ponto inconsistente com o Google Patents, sinalizado neste estudo, foi o fato de que ele não indexa os registros do escritório brasileiro, apesar de a base apresentar uma cobertura documental muito superior às demais bases de dados estudadas.

Diante do que foi exposto, percebe-se que a estrutura política de disponibilização e a estruturação dos dados — no caso, os registros de patentes — variam conforme as instituições responsáveis pela administração das fontes de informações consultadas. Desta forma, pode-se concluir que, nos estudos de patentes, a escolha da fonte de dados é muito importante para a qualidade e veracidade do estudo que será desenvolvido. Diante disso, deve-se escolher a fonte de informação conforme uma série de critérios que tenha relação direta com o objeto a ser estudado. Ou seja, é preciso que os estudos de natureza patentométrica levem em consideração a abrangência da base, a pontualidade com que os registros são inseridos e as possibilidades de recuperação e exportação dos registros para as análises dos dados. Isso porque esses critérios influenciam diretamente na construção e inferência dos estudos sobre o objeto analisado. No presente caso, cada base de dados utilizada para verificar a produção intelectual da Universidade apresentou um valor e significado distinto, o que faz toda a diferença dentro de um processo de tomada de decisão.

Referências

ARAÚJO, Vânia Maria Rodrigues Hermes. A patente como ferramenta da informação. **Ciência da Informação**, v. 10, n. 2, p. 27-32, 1981. Disponível em: <http://revista.ibict.br/ciinf/article/view/148> Acesso em: 31 ago. 2019.

BRASIL. **Decreto nº 8.777, de 11 de maio de 2016**. Institui a Política de Dados Abertos do Poder Executivo federal. Brasília: Presidência da República, 2016. Disponível em: http://www.planalto.gov.br/ccivil_03/_Ato2015-2018/2016/Decreto/D8777.htm. Acesso em: 31 ago. 2018.

- BRASIL. Ministério da Educação. Instituto Nacional de Estudos e Pesquisas Educacionais Anísio Teixeira. **Índice Geral de Cursos (IGC)**. Brasília: MEC, 2018. Disponível em: <http://portal.inep.gov.br/indice-geral-de-cursos-igc->. Acesso em: 09 set. 2018.
- BRASIL. Ministério do Planejamento, Desenvolvimento e Gestão. Secretaria de Tecnologia da Informação. **RPI de Patentes (a partir de 31/Jan/2017)**. Brasília: Portal Brasileiro de Dados Abertos, 2017.
- CALDAS, Max Silva.; SILVA, Emanuel Costa Claudino. Fundamentos e aplicação do Big Data: como tratar informações em uma sociedade de yottabytes. **Revista Bibliotecas Universitárias: Pesquisa, Extensão e Perspectiva**, v. 3, n. 1, p. 65-85, jan./jun. 2016. Disponível em: <https://periodicos.ufmg.br/index.php/revistarbu/article/view/3086> Acesso em: 31 ago. 2019.
- CLARIVATE ANALYTICS. **Derwent World Patents Index**. [S.l.]: Clarivate Analytics, 2017. Disponível em: <https://clarivate.com/derwent/> Acesso em: 31 ago. 2019.
- CONSONI, Letícia Angheben El Ammar. **Produção tecnológica em biodiesel**: análise das características dos depósitos de patentes indexadas na Derwent Innovations Index entre 1983 e 2015. 2017. 193 f. Dissertação (Mestrado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informações, Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2017. Disponível em: <https://lume.ufrgs.br/handle/10183/157680> Acesso em: 31 ago. 2019.
- CONSONI, Letícia Angheben El Ammar; MOURA, Ana Maria Mielniczuk de. Patentes de biodiesel indexadas na derwent innovations index entre 2008 e 2009: primeiras considerações. *In*: ENCONTRO NACIONAL DE PESQUISA EM CIÊNCIA DA INFORMAÇÃO, 17., Salvador, 2016. **Anais [...]** Salvador: ANCIB, 2016.
- CUNHA, Murilo Bastos. **Para saber mais**: fontes de informação em ciência e tecnologia. Brasília: Briquet de Lemos, 2001.
- FRANÇA, Ricardo Orlandi. A patente. *In*: CAMPELLO, Bernadete Santos; CENDÓN, Beatriz Valadares; KREMER, Jeannette Marguerite (Org.). **Fontes de informação para pesquisadores e profissionais**. Belo Horizonte: UFMG, 2000. cap. 12, p. 152-182.
- GARCIA, Joana Coeli Ribeiro. Os paradoxos da patente. **DataGramaZero**, v. 7, n. 5, p. A04-0, 2006. Disponível em: <http://www.brapci.inf.br/v/a/4129>. Acesso em: 20 jun. 2018.
- GHESTI, Greice Ferreira (coord.). **Tutorial de busca nos principais bancos de patentes**. Brasília: Centro de Apoio ao Desenvolvimento Tecnológico da Universidade de Brasília, 2016. Disponível em: <http://www.cdt.unb.br/pdf/programaseprojetos/nupitec/LIVRO-BANCO%20DE%20PATENTES.compressed.pdf> Acesso em: 31 ago. 2019.
- GOOGLE. **About Google Patents**. [S.l.]: Google, c2018. Disponível em: <https://support.google.com/faqs/answer/6390996..> Acesso em: 04 jan. 2018.
- JAGHER, Tatiane. **Busca em banco de dados de patentes**. Curitiba: Agência de Inovação UTFPR, [2010?]. Disponível em: <http://www.utfpr.edu.br/medianeira/estrutura/diretorias/direc/downloads/PROCEDIMENTOPARAPESQUISAUMAPATENTE.pdf>. Acesso em: 09 set. 2018.
- MAGNUS, Ana Paula Medeiros. **Produção tecnológica e científica**: panorama das patentes e artigos dos pesquisadores dos Programas de Pós-Graduação do Instituto de Química da UFRGS. 2018. 156 f. Dissertação (Mestrado em Comunicação e Informação) – Programa de Pós-Graduação em Comunicação e Informação, Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2018.
- MARICATO, João de Melo; NORONHA, Daisy Pires. Coclasseificação em artigos e patentes em biodiesel. **Liinc em Revista**, v. 9, n. 1, p. 85-102, 2013. Disponível em: <http://revista.ibict.br/liinc/article/view/3386> Acesso em: 31 ago. 2019.
- MONTEIRO, Silvana Drummond; FERNANDES, Rogério Paulo Muller; DECARLI, Gian Carlo; TREVISAN, Gustavo Lunardelli. Sistemas de recuperação da informação e o conceito de relevância nos mecanismos de busca: semântica e significação. **Encontros Bibli**, v. 22, n. 50, p. 161-175, 2017. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2017v22n50p161> Acesso em: 31 ago. 2019.
- MOURA, Ana Maria Mielniczuk de; ROZADOS, Beatriz Frota; CAREGNATO, Sônia Elisa. Interações entre ciência e tecnologia: análise da produção intelectual dos pesquisadores-inventores da primeira carta-patente da UFRGS. **Encontros Bibli**, v. 11, n. 22, p. 1-15, 2006. Disponível em: <https://periodicos.ufsc.br/index.php/eb/article/view/1518-2924.2006v11n22p1> Acesso em: 31 ago. 2019.

MUELLER, Suzana Pinheiro Machado; PERUCCHI, Valmira. Universidades e a produção de patentes: tópicos de interesse para o estudioso da informação tecnológica. **Perspectivas em Ciência da Informação**, v. 19, n. 2, p. 15-36, abr./jun. 2014. Disponível em: <http://www.scielo.br/pdf/pci/v19n2/03.pdf> Acesso em: 31 ago. 2019.

OKUBO, Yoshiko. **Bibliometric indicators and analysis of research systems**: methods and examples. Paris: OECD, 1997. (OECD Science, Technology and Industry Working Papers, 1997/01). Disponível em: https://www.oecd-ilibrary.org/science-and-technology/bibliometric-indicators-and-analysis-of-research-systems_208277770603 Acesso em: 31 ago. 2019.

SOUTO, Leonardo Fernandes. Recuperação de informações em bases de dados: usos de tesouro. **Transinformação**, v. 15, n. 1, p. 73-81, jan./abr. 2003. Disponível em: http://www.scielo.br/scielo.php?pid=S0103-37862003000100006&script=sci_abstract&tlng=pt Acesso em: 31 ago. 2019.

SOUZA, Renato Rocha. Sistemas de recuperação de informações e mecanismos de busca na web: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v. 11, n. 2, p. 161-173, maio/ago. 2006. Disponível em: <http://portaldeperiodicos.eci.ufmg.br/index.php/pci/article/view/320> Acesso em: 31 ago. 2019.

SPINAK, Ernesto. **Indicadores cientométricos de patentes**: aplicaciones y limitaciones. Madrid: [s.n.], mar. 2003. Disponível em: http://www.riicyt.edu.ar/interior/normalizacion/III_bib/Spinak.pdf. Acesso em: 01 ago. 2018.

WETHERBEE, Ian. **Google Patents Public Datasets**: connecting public, paid, and private patent data. [S.l.]: Google, 2017. Disponível em: <https://cloud.google.com/blog/products/gcp/google-patents-public-datasets-connecting-public-paid-and-private-patent-data>. Acesso em: 09 set. 2018.

WORLD INTELLECTUAL PROPERTY ORGANIZATION. **PatentScope**. [S.l.]: PatentScope, 2018. Disponível em: <https://patentscope.wipo.int/search/pt/search.jsf>. Acesso em: 31 ago. 2018.

Artigo submetido em: 22/10/2019.

Aceito em: 29/12/2019.

UFCA UNIVERSIDADE
FEDERAL DO CARIRI

Centro de Ciências Sociais Aplicadas
Mestrado Profissional em Biblioteconomia



Este periódico é uma publicação do [Programa de Pós-Graduação em Biblioteconomia](#) da [Universidade Federal do Cariri](#) em formato digital e periodicidade semestral.