

Busca, recuperação e padronização de metadados de patentes: limitações e dificuldades

Autor: Thiago Monteiro Alves

Orientador: Ana Maria Mielniczuk de Moura

1 INTRODUÇÃO

Este estudo visa demonstrar as dificuldades e limitações encontradas na busca, recuperação e padronização de metadados de patentes coletadas na base de dados Derwent Innovations Index (DII), de propriedade da empresa Clarivate Analytics e que pode ser acessada, mediante assinatura, pelo Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), utilizada em estudos métricos (MARICATO, 2010; MOURA, 2009).

2 METODOLOGIA

É um estudo de natureza básica, abordagem quali-quantitativa, com objetivo descritivo e utiliza-se de procedimentos patentométricos para contemplar os objetivos do trabalho. Serviu como base para a pesquisa “Produção tecnológica e científica: panorama das patentes e artigos dos pesquisadores dos programas de pós-graduação do Instituto de Química da UFRGS”, projeto de mestrado da aluna Ana Paula Medeiros Magnus, vinculada ao projeto “Interação entre Ciência e Tecnologia em artigos e patentes no Brasil”, coordenado pela Prof^a. Dr.^a Ana Maria Mielniczuk de Moura. Este estudo foi fruto do auxílio prestado pelo autor na busca, recuperação e padronização de dados para a pesquisa citada. A coleta dos dados foi realizada entre os meses de dezembro de 2017 e janeiro de 2018. Os registros foram recuperados a partir de estratégias de busca elaboradas com as diversas variações possíveis do nome e sobrenome dos pesquisadores estudados. Para a confirmação dos dados coletados na DII, foi realizado uma conferência dos registros obtidos através do Instituto Nacional de Propriedade Industrial (INPI) e do Currículo Lattes de cada pesquisador, para constatar a veracidade das informações recuperadas. A amostra foi composta pelos 123 pesquisadores que possuíam patentes indexadas na DII, totalizando 271 documentos de patentes.

3 RESULTADOS

A coleta foi feita individualmente para cada pesquisador. A primeira dificuldade encontrada foi pela falta de padronização dos nomes dos inventores (Quadro 1).

Quadro 1 – Forma do nome dos inventores encontrado na DII

Nome	Primeira Forma Pesquisada	Segunda Forma Pesquisada	Formas Encontradas nos Registros da DII
Carlo Requião da Cunha	Cunha, C R	Requião da Cunha C OR Da Cunha C R OR Da Cunha C. R. OR Cunha Carlo R. da OR Da Cunha Carlo R. OR da Cunha C.R. OR Da Cunha Carlo	1. REQUIAO DA CUNHA C 2. DA CUNHA C R 3. CUNHA C R D

Fonte: dados da Pesquisa.

A segunda dificuldade encontrada, foi tentar confirmar se realmente os registros recuperados pertenciam aos pesquisadores que fizeram parte do *corpus* da pesquisa. Realizando a busca no INPI e no Currículo Lattes, foi possível confirmar se os registros recuperados na DII pertenciam (ou não) aos mesmos. No INPI foram recuperados 251 registros, já no Currículo Lattes esse número chegou a 453.

A terceira dificuldade encontrada ocorreu pela diferença da forma como os campos de busca são registrados tanto na DII como no INPI (Quadro 2).

Quadro 2 – Diferença da forma de registro nos campos da DII e INPI

	Como é registrado na DII	Como é registrado no INPI
Número da Patente	Campo PN = BR102014007606	Campo Nº do Pedido = BR 10 2014 007606 9
Código de Classificação (CIP)	Campo IP = C07D-233/54	Campo Classificação IPC = C07D 233/54
Data de Depósito da Patente	Campo PD = 28 Mar 2014	Campo Data Depósito = 28/03/2014
Nome do Inventor	Campo AU = (RUSSOWSKY D)	Campo Nome Inventor = Dennis AND Russowsky
Nome do Depositantes	Campo AE = UNIV FEDERAL DO RIO GRANDE DO SUL (UYRI-Non-standard)	Campo Nome Depositante/Titular = UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL (BR/RS)

Fonte: dados da Pesquisa.

4 CONSIDERAÇÕES FINAIS

Podemos constatar que a DII não indexa todas as patentes publicadas pelo INPI, demonstrando inconsistência das informações contidas na base. Além disto, há uma grande dificuldade de confirmar os registros de autoria das patentes pois a base não possui uma padronização que permita discernir fielmente os autores. Percebe-se por fim, a importância e necessidade de se estudar profundamente as bases de dados que publicam patentes, a fim de compreender suas coberturas, formas de indexação, recuperação, apresentação e padronização dos dados para que possam ser utilizadas como ferramentas em pesquisas acadêmicas.

REFERÊNCIAS

- MARICATO, J. M. **Dinâmica das relações entre ciência e tecnologia**: estudo bibliométrico e cientométrico de múltiplos indicadores de artigos e patentes em biodiesel. 2010. 378 f. Tese (Doutorado em Ciência da Informação) – Escola de Comunicação e Artes, Universidade de São Paulo, São Paulo, 2010.
- MOURA, A. M. M. **A interação entre artigos e patentes**: um estudo cientométrico da comunicação científica e tecnológica em biotecnologia. 2009. 270 f. Tese (Doutorado em Comunicação e Informação) – Faculdade de Biblioteconomia e Comunicação, Universidade Federal do Rio Grande do Sul, Porto Alegre, 2009.