

ESTUDO SISTEMATIZADO PARA SELEÇÃO DE TERMOS DE BUSCA PARA PROPAGAÇÃO VEGETATIVA DE CANA-DE-AÇÚCAR

Cecilia Anita Hasner Domjan¹; Eduardo Winter²; Ricardo Carvalho Rodrigues^{3*}

^{1,2,3}Instituto Nacional da Propriedade Industrial, Academia de Propriedade Intelectual e Inovação, Rio de Janeiro, RJ, Brasil.

Rec.: 18.09.2016 Ace.: 22.12.2016

RESUMO

As técnicas de propagação de cana-de-açúcar nos últimos anos vêm direcionando seus esforços de pesquisa para ampliar a produção de mudas, os quais são registrados em documentos científicos e técnicos. Entretanto, não existe uma metodologia sistematizada para a recuperação destas informações que possam subsidiar pesquisas futuras. Neste sentido, o foco deste trabalho é estudar a utilização de termos de busca para o levantamento bibliográfico sobre tecnologias de propagação vegetativa da cana-de-açúcar e espécies similares de forma sistemática. A metodologia consiste em oito etapas, utilizando o Agrovoc para levantamento de palavras chaves e técnicas de avaliação da efetividade dos termos nos documentos recuperados. Os termos “*plant*” e “*seed*” mostraram-se mais precisos que “*vegetative*” e “*propagation*” nas bases da Web of Science, Scopus e Derwent Innovation Index, com maior destaque para a Scopus. Estudos futuros permitirão compreender se estes resultados refletem o estado de desenvolvimento tecnológico da propagação vegetativa da cana-de-açúcar.

Palavras-chave: Termos de busca. Artigos. Patentes.

SYSTEMATIZED STUDY FOR SEARCH TERMS SELECTION FOR VEGETATIVE PROPAGATION OF SUGARCANE

ABSTRACT

The sugarcane propagation techniques have been focusing their research efforts to expand the production of seedlings, which are recorded in scientific and technical documents. However, there is no systematic methodology for the information recovery that can support future research. In this sense, the goal of this study is analyses the use of search terms for the literature on vegetative propagation technologies of sugarcane and similar species systematically. The methodology is based on eight stages using the AGROVOC for identifying keywords and evaluation techniques of the effectiveness of the terms in the retrieved documents. The results showed that the terms “*plant*” and “*seed*” are more accurate than “*vegetative*” and “*propagation*” in the databases Web of Science, Scopus and Derwent Innovation Index, highlighting Scopus. Future studies will help to understand if these results reflect the state of technological development of vegetative propagation of sugarcane.

KEYWORDS: Search terms. Papers. Patents.

Área Tecnológica: Ciência da Informação e Agricultura

*Autor para correspondência: ricardo.rodrigues@inpi.gov.br

INTRODUÇÃO

O setor canavieiro passou por uma modernização muito rápida e de forma desigual nos últimos anos, onde a mecanização da colheita saltou de 28 para 89% na região Centro-Sul no período de 2000 a 2013, impulsionada pela regulamentação da Lei do estado de São Paulo, n.º 11.241/2002, que dispõe sobre a eliminação gradativa da queima da palha da cana-de-açúcar e dá outras providências (NEVES *et al.*, 2014). Neste processo de modernização, destacam-se avanços tecnológicos associados a propagação de cana-de-açúcar, melhoramento genético e expansão dos canaviais comerciais em novas fronteiras agrícolas, esforços voltados para aumentar a produtividade e reduzir custos (CARVALHO & FURTADO, 2013). Em relação as técnicas de propagação de cana-de-açúcar, os esforços em pesquisa e desenvolvimento (P&D) são voltados para ampliar a produção de mudas e gerar mudas mais eficientes nos sistemas mecanizados. A forma tradicional de plantio de cana é realizada de forma manual, utilizando cortes de colmos¹ em seções e depositados horizontalmente em sulcos (CTC – CENTRO DE TECNOLOGIA CANAVIEIRA., 2015). Posteriormente, foram desenvolvidas técnicas de propagação utilizando mudas pré-brotadas a partir de colmos, mudas a partir da gema lateral ou utilizando técnicas de micropropagação a partir de tecido meristemático, e mais recentemente, a produção de sementes artificiais ou encapsulamento de embriões somáticos em estágio avançado de desenvolvimento em tubetes especiais (NIEVES *et al.*, 2003; CTC – CENTRO DE TECNOLOGIA CANAVIEIRA, 2015). Apesar destes avanços, Nieves *et al* (2003) relatam a dificuldade em aceitar comercialmente as sementes artificiais, apesar das vantagens frente as técnicas tradicionais, tais como não precisar tanto de mão de obra, nem grandes áreas para viveiros ou logística de armazenamento e transporte, irrigação e aclimatização. A modo de exemplo, pode-se citar o caso da Syngenta, que lançou em 2010 a PLENE®, uma muda pré-germinada que prometia aumentar a produtividade entre 5 e 10%, porém um ano depois, a empresa precisou retirar o produto do mercado por falhas na produção industrial em grandes volumes, acarretando um prejuízo de milhões de dólares (VALOR, 2014). Em 2014 a Syngenta firmou um acordo de licenciamento com a canadense New Energy Farms para produzir gemas encapsuladas de cana-de-açúcar para plantio no Brasil, uma tecnologia denominada CEEDS, que pretende ser lançado na safra de 2016/2017 (SYNGENTA, 2014).

Outros esforços em P&D em empresas e centros de pesquisa merecem destaque. A Basf também inovou nesta área lançando em 2013 a Agmusa, uma muda pré-brotada (MPB) que permite um ganho de produtividade de 20 a 40% (VALOR, 2014). A sua vez, o Programa Cana do Instituto Agrícola de Campinas - IAC também desenvolveu uma MPB de cana-de-açúcar e seus métodos de utilização em viveiros tradicionais, sendo possível utilizá-las tanto em plantio manual como mecanizado, permitindo corrigir falhas nas fileiras de cana (LANDELL *et al.*, 2012). O CTC – Centro de Tecnologia Canavieira vem desenvolvendo pesquisas nesta área desde 2008 e já possui uma Biofábrica para a produção de “sementes” de cana a partir do meristema, que será lançada entre 2017 e 2018 (VALOR, 2014).

Todos estes esforços de P&D são registrados em documentos científicos e tecnológicos, sejam através das publicações em revistas científicas ou através de documentos de patente de invenção. Esta interação entre a Ciência e a Tecnologia tem sido abordada na área de biotecnologia no Brasil utilizando a Cientometria como metodologia de análise, dedicada a compreender os fluxos do conhecimento entre os inventores e autores que transitam em ambas esferas (MOURA, 2009).

¹ Tipo de caule ou haste de gramíneas em que nós e entrenós são bem visíveis. O termo "nó" indica a parte do colmo a partir do qual cresce a parte aérea e raízes; cada planta contém muitos nós. Na posição de cada nó, forma-se a gema que pode desenvolver em planta inteira. O material apropriado para corte são pedaços de cana cortados de plantas saudáveis com oito a quatorze meses de idade e os cortes são retirados de plantas que cresceram geralmente a partir de cortes (CTC, 2014)

Estudos sobre a produção científica e tecnológica também serviram de indicadores para analisar o sistema nacional de inovação e sua distribuição espacial no país (ALBUQUERQUE *et al.*, 2002), bem como a análise de correlação das classificações (coclassificação) de artigos e patentes em diferentes tipos de organizações (públicas e privadas) na área de biodiesel (MARICATO & NORONHA, 2013). Winter *et al.* (2010) utilizaram a análise de patentes no mapeamento da cadeia produtiva do etanol da década de 70 até 2010, onde observaram que as principais tecnologias protegidas no país se referem a equipamentos e processos de plantio, colheita e preparo do solo, e não nas áreas relacionadas com biotecnologia. O campo da biotecnologia engloba técnicas de fermentação, engenharia genética e também técnicas de propagação vegetativa por cultivo de tecidos, foco do presente estudo.

Os artigos e patentes apresentam algumas divergências básicas na maneira em que o trabalho ou invenção são abordados, dado pelo maior grau de liberdade apresentado pela publicação, em comparação à maior objetividade de um pedido de patente (DIAS & ALMEIDA, 2013). Ademais, Dias e Almeida (2013) destacam que:

“A produção científica nem sempre parte do pressuposto de que existe um problema a ser resolvido e sim que existe algo a ser revelado que até agora não foi. A revelação de algo ainda não descrito gera novas informações, que são a base para novas interpretações, fechando um círculo virtuoso que, em grande parte, contribui para a geração de novos conhecimentos. No entanto, o simples fato de revelar algo novo não é suficiente para que uma publicação se torne um pedido de patente”. (DIAS & ALMEIDA, 2013, p.5).

No mundo acadêmico, a fonte de informação mais utilizada é o artigo científico, constituindo-se em um veículo de transmissão do conhecimento produzido pelos pesquisadores e servindo de literatura-base para corroborar tanto os estudos já existentes como subsidiar novas pesquisas (SILVEIRA, 2005). Não obstante, o uso de documentos de patente como fonte de informação por parte de pesquisadores é muito pouco explorada no país, das 586 teses e dissertações da Unicamp analisadas no período de 2000 a 2007, 16,4% delas tiveram patentes citadas (RAVASCHIO *et al.*, 2010). Já a informação tecnológica contida em documentos de patente tem sido utilizada para diferentes fins, tais como estudo de tendências tecnológicas, estudo de mercado, busca de anterioridade para conhecer o estado da técnica, subsídio a P&D, tomada de decisão empresarial, entre outros (FERREIRA *et al.* 2009, OLIVEIRA *et al.* 2005, BONINO *et al.* 2010). Atribui-se uma grande vantagem das patentes frente aos artigos científicos, não só pela riqueza da informação tecnológica descrita nos documentos, mas pela facilidade de recuperar documentos pela universalidade dos códigos de indexação nos dados bibliográficos (códigos INID²). Estes códigos identificam todas as informações que constam da primeira página ou folha de rosto, tendo indicações de números, datas, origem, propriedade e técnica.

A recuperação de documentos de patente depende de uma boa estratégia de busca, a qual utiliza tanto palavras-chaves, como classificação de patentes por campo tecnológico específico. A sua vez, a elaboração da estratégia de busca requer não só um domínio de conhecimento prévio sobre o tema que se pretende analisar, como também conhecer as ferramentas de busca e as bases de dados de patentes (MONCHAUX *et al.*, 2015). Desta forma, tem-se que um dos problemas associados a escolha de uma estratégia de busca eficiente radica em levantar os termos (palavras-chave) e as classificações internacionais de patentes (IPC, CPC, FI, US Class³). Na literatura é possível

² INID é o acrônimo de “*Internationally agreed Numbers for the Identification of (bibliographic) Data*”.

³ IPC – International Patent Classification; CPC – Cooperative Patent Classification; FI – é uma classificação de patentes utilizada pelo Escritório de Patentes do Japão (JPO); US Class – classificação norte-americana de patentes.

Domjan, C.A.H; Winter, E.; Rodrigues, R.C. Estudo sistematizado para seleção de termos de busca para propagação vegetativa de cana-de-açúcar.

encontrar metodologias específicas para identificar as palavras-chave mais adequadas (NIJHOF 2007, MOELLER & MOEHRE 2015, ALBERTS 2011), porém muitas delas requerem o uso de ferramentas para ampliação dos termos ou verificação dos critérios de cobertura (*recall*) e precisão (*precision*) dos documentos recuperados⁴. Outro problema associado à recuperação de documentos é a análise destes, pois não necessariamente os documentos mais relevantes encontram-se no topo da lista dos documentos recuperados. Neste sentido, diversos autores recomendam metodologias para a priorização de documentos de patentes através de ranking dos resultados descritos na literatura (ROBERTSON 1977, MATERNE & SLEIGHTHOLME, 2014, GUO & GOMES 2009, MAGDY & JONES 2010, CAO et al 2013) e em metodologias protegidas por patentes ((KOBE UNIV & KANEKA CORP 2014, ZMC INTELLECTUAL DATA SCIENCE & TECHNOLOGY CO. LTD 2011, GOEDKEN 2005). Entre as técnicas descritas sobre priorização destacam-se as seguintes:

- i) Ranking de Frequência ou Ocorrência dos termos utilizados – nesta técnica é analisado o número de ocorrência de cada palavra-chave ou da classificação de patentes utilizada na estratégia de busca em cada documento, sendo considerado o mais relevante o documento com o número mais elevado (MATERNE & SLEIGHTHOLME, 2014);
- ii) *Horváth Ranking* ou *Facet ranking* – cada palavra-chave ou classificação de patente empregada na estratégia de busca é considerada uma “faceta” e se analisa quantas facetes são utilizadas em cada documento de patente recuperado. O documento que possuir o maior número de facetes é considerado o mais relevante (MATERNE & SLEIGHTHOLME, 2014);
- iii) *Ranking pelo tamanho de família de patentes* – A priorização dos documentos de patentes leva em consideração o tamanho da família de patente, no qual quanto maior a família de patente de um documento, maior a sua relevância (ZMC INTELLECTUAL DATA SCIENCE & TECHNOLOGY CO. LTD, 2011);
- iv) *Cosine relevance ranking* – descrito por Salton & Buckley (1988), refere-se a uma metodologia que atribui peso aos termos (palavras-chave e classificações de patentes) utilizados na estratégia de busca. A técnica contempla três vetores: o vetor de frequência dos termos nos documentos recuperados, o vetor da frequência dos termos na coleção (banco de dados) e o vetor de normalização cosine ou o comprimento do vetor de ambas variáveis, a qual mede a semelhança entre elas;
- v) *Horváth-Materne* ou *Pivot ranking* – consiste em reorganizar os resultados levando em consideração alguns critérios de tal forma que os mais relevantes apareçam no topo da lista e foi desenvolvida e implementada pelo Escritório Europeu - EPO (MATERNE & SLEIGHTHOLME, 2014). Esta técnica se baseia em 3 passos fundamentais: 1) busca de palavras-chave no texto completo do documento de patente; 2) aplicação do *Facet Ranking* no resumo dos documentos recuperados; 3) aplicação do ranking de frequência para cada set das facetes (MATERNE & SLEIGHTHOLME, 2014).

Tanto a técnica de Ranking de Frequência ou Ocorrência dos termos utilizados como a técnica *Facet Ranking* tem a vantagem na forma simples de serem aplicadas, pois não requerem o emprego de fórmulas matemáticas e análise de dados elaborados, dispensando assim o uso de programas de computador.

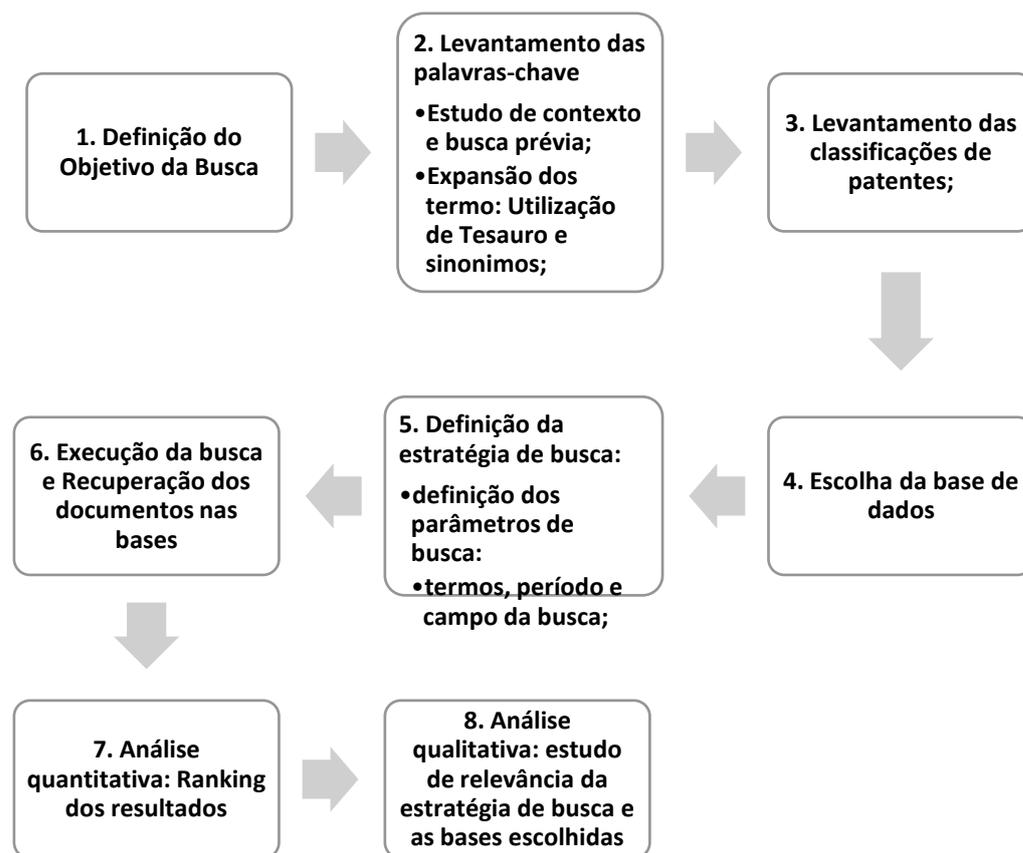
⁴ *Recall* e *Precision* são os principais parâmetros de avaliação da qualidade da estratégia de busca (ZIMAN, 1969). *Recall* é definido como sendo a proporção de documentos relevantes em uma coleção ou base que são recuperados em uma determinada pesquisa e o grau de precisão é definida como a proporção de documentos relevantes consideradas pertinentes para a estratégia de busca, ou a relevância dos documentos recuperados (VAN DER DRIFT, 1991)

Levando em consideração a importância do uso da informação, científica e tecnológica, para análise dos avanços tecnológicos da propagação vegetativa de cana-de-açúcar, buscou-se identificar uma metodologia sistematizada de pesquisa, que permitisse avaliar sinergias entre inventores e pesquisadores. Tal metodologia não foi encontrada, nem mesmo uma que pudesse ser replicada para pesquisa da mesma família da cana-de-açúcar, a família Poaceae. Neste contexto, o objetivo deste trabalho é estudar a utilização de diferentes termos de busca para o levantamento bibliográfico que englobe publicações científicas e tecnológicas (documentos de patente) da cana-de-açúcar e outras espécies da mesma família de forma sistemática e que sirvam de referência para pesquisas no setor agroindustrial.

METODOLOGIA

O estudo do emprego dos termos de busca para levantamento de informações científicas e técnicas foi realizado em oito etapas e foi baseado na metodologia de recuperação da informação de patentes descrito pelos autores Hunt, Nguyen e Rodgers (2007), conforme mostra o Quadro 1.

Quadro 1 - Etapas da Metodologia Sistematizada para levantamento bibliográfico de literatura patentária e científica sobre Mudanças de Cana-de-Açúcar.



Fonte elaboração própria baseado em Hunt, Nguyen e Rodgers (2007).

A primeira etapa consiste em definir claramente o objetivo da busca. No caso de avanços tecnológicos sobre propagação vegetativa de cana-de-açúcar, é importante considerar espécies da

mesma família Poaceae que possuem interesse comercial, servindo de benchmarking, tais como o sorgo e o bambu.

A segunda etapa se refere ao levantamento das palavras-chave. Desta forma, realizou-se um levantamento dos conceitos referente a propagação vegetativa utilizando um tesouro específico para a área tecnológica, o AGROVOC Multilingual agricultural⁵. Os principais conceitos foram:

- “seedling” ou “germination” para produção de plântulas,
- “seedlings” ou “plantlets” para plântulas,
- “vegetative propagation” para propagação vegetativa e
- “micropropagation” para técnicas de cultivo de tecidos meristemáticos ou micropropagação.

A terceira etapa refere-se ao levantamento das classificações de patentes. Para tal, utilizou-se a publicação oficial de Classificação Internacional de Patentes (CIP) através da busca por termos e foi identificada a seguinte classificação:

A01H - Novas Plantas ou Processos para Obtenção das Mesmas; Reprodução de Plantas por Meio de Técnicas de Cultura de Tecidos;

A escolha das bases foi realizada na etapa 4. Para a literatura não-patentária foram escolhidas as bases Web of Science, da Thomson Reuters, e a Scopus, da Elsevier, devido a abrangência de periódicos indexados. A base da Scopus difere da Web of Science, pois cobre mais periódicos da América Latina e o Caribe. Para a pesquisa da literatura patentária foi escolhida unicamente a base do Derwent Innovation Index, da Thomson Reuters, que abrange mais de 14,3 milhões de invenções básicas de 40 escritórios de patentes em todo o mundo.

A estratégia de busca (etapa 5) consiste em utilizar o máximo de palavras-chave da tecnologia e os principais sinônimos, delimitando o período de tempo que se pretende levantar os documentos.

A partir da execução da busca (etapa 6), os documentos recuperados são exportados para uma planilha Excel, separadas pelo ano da publicação, autores ou inventores, título e resumo.

A etapa 7 consiste na análise quantitativa dos dados recuperados, ou seja, técnicas capazes avaliar a efetividade da utilização dos termos empregados na estratégia. Utilizaram-se a metodologia de priorização por frequência de palavras e *Facet Ranking*.

A última etapa consiste na análise qualitativa, realizada através da leitura dos documentos que possuem maior ranking, verificando a pertinência do documento em relação a propagação vegetativa das espécies de cana, sorgo e bambu.

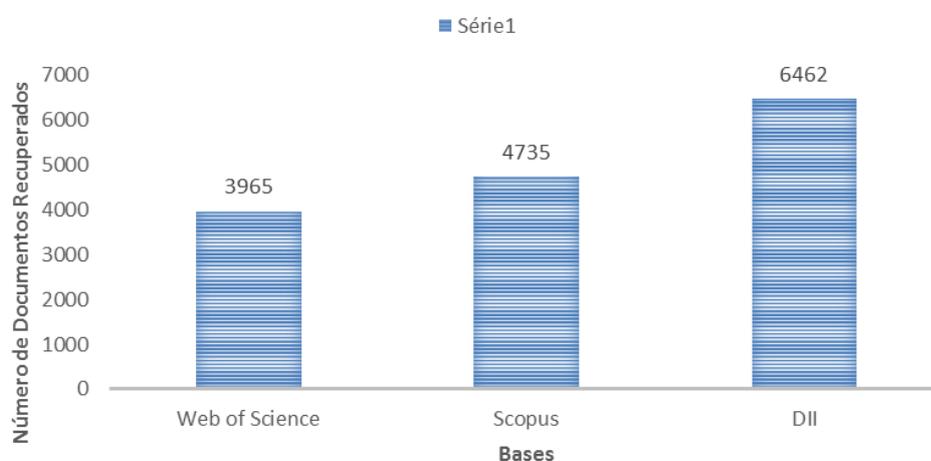
RESULTADOS E DISCUSSÃO

O número de documentos recuperados utilizando as palavras-chave “sugarcane or saccharum”, “bambu* or bamboo*”, “sorghum or sorgo”, “vegetative or plant*” and “propagation or production” ou “seed* or germination or plantlets” no período de 2005 a 2014 foram de 3.965 na base Web of Science, 4.735 na base da Scopus e 6.462 na base de patentes (Figura 1). Percebe-se que o volume de documentos recuperados na base do Scopus é maior que a Web of Science, o qual

⁵ O thesaurus AGROVOC, uma junção da agricultura e vocabulário, é um vocabulário controlado multilíngue abrangendo todas as áreas de interesse da Organização para a Alimentação e Agricultura das Nações Unidas (FAO), incluindo alimentação, nutrição, agricultura, pesca, silvicultura e meio ambiente. Encontra-se disponível no site: <http://aims.fao.org/standards/agrovoc/functionalities/search>.

pode ser atribuído as próprias características da base, que contempla países da América Latina e Caribe, mostrando assim o interesse desta região em P&D da cana-de-açúcar, principalmente se for levado em consideração que o Brasil é o líder mundial em produção de cana (FAO, 2014). A base de patentes do Derwent Innovation Index apresentou o maior número de documentos recuperados, o qual reflete não só a abrangência da cobertura da base, como reforça a importância dos documentos de patentes frente aos demais documentos devido as características da estrutura do documento de patente que facilitam sua indexação nas bases de dados. Outra interpretação possível pode ser atribuída ao interesse dos países em proteger as tecnologias relacionadas com a propagação vegetativa.

Figura 1 - Número de documentos recuperados nas diferentes bases utilizando termos associados com a propagação vegetativa de cana-de-açúcar, sorgo e bambu no período de 2005 a 2014.



Fonte: Elaboração própria em base à pesquisa feita na Web of Science, Scopus e Derwent Innovation Index (DII).

A partir do conjunto de documentos recuperados, foram selecionados uma amostra para análise quantitativa e qualitativa. Esta amostra foi composta pelos primeiros 500 documentos priorizados por relevância em suas respectivas bases e exportadas para planilha Excel. O resultado da análise quantitativa, medido pelo número de ocorrência dos termos utilizados na estratégia de busca em cada base encontra-se detalhada no Quadro 2. Percebe-se uma diferença entre as bases, principalmente entre as bases de artigos científicos e de patentes. A base Scopus apresentou maior ocorrência com termos relacionados a cana-de-açúcar em comparação as demais bases. Pode-se inferir que esta diferença se deve à indexação de revistas latino-americanas, onde o cultivo da cana é mais relevante. As palavras-chave relacionadas com a propagação vegetativa mais frequentes foram “*plant*” e “*seed*”, porém a primeira encontra-se em maior volume nas bases de artigos, enquanto que a segunda é mais expressiva na base de patentes. Esta diferença já mostra a objetividade dos documentos de patentes frente a artigos, pois “*seed*” abrange tanto sementes como mudas. As palavras-chave relacionadas com espécies comerciais da família Poaceae mais frequentes nas bases de artigos foram “*sugarcane*”, “*sorghum*”, enquanto que na base de patentes o truncamento da palavra bambu teve maior ocorrência. O termo “*sacharum*”, nome científico da cana-de-açúcar, só mostrou um número expressivo na base do Scopus. Estes resultados demonstram a importância de avaliar previamente a utilização de termos nas bases que gerem maior precisão na recuperação da informação.

Quadro 2: Número de Ocorrências registradas e Porcentagem para cada palavra chave utilizada na estratégia de busca nas diferentes bases do Portal Capes. Foram analisados 500 documentos de cada base.

| Palavras chaves | Número de Ocorrência nas Bases | | |
|---|--------------------------------|--------|-----|
| | WoS | Scopus | DII |
| Palavras chaves relacionadas com a família Poaceae | | | |
| Sorghum | 203 | 121 | 129 |
| Sugarcane | 181 | 378 | 86 |
| Bamb | 50 | 4 | 278 |
| Saccharum | 28 | 174 | 3 |
| Palavras chaves relacionados com propagação vegetativa | | | |
| Plant | 365 | 399 | 191 |
| Seed | 185 | 170 | 437 |
| Germination | 53 | 56 | 21 |
| Vegetative | 20 | 29 | 2 |
| Propagation | 15 | 34 | 7 |
| Plantlets | 11 | 62 | 1 |

Fonte: Elaboração própria.

A partir destes resultados, aplicou-se a metodologia de priorização por *Facet Ranking*, utilizando as seguintes palavras-chave no campo resumo: plant, plantlets, vegetative, germination, saccharum, seed, propagation, sugarcane, bamb e sorghum (Quadro 3). O resultado mostrou que o número máximo de palavras chave presente em cada documento, doravante denominado facetas, foi de 6 palavras-chave no resumo do documento. O número zero do ranking pode ser interpretado como nenhuma palavra-chave da estratégia estava presente no resumo, porém o documento foi recuperado levando em consideração outros campos do documento, tal como o título ou no texto completo do documento. A maior frequência ocorreu com duas e três facetas (Quadro 3), ou seja, cada documento recuperado contempla a combinação de duas a três palavras-chaves no resumo. Esta informação, aliada a uma análise dos termos utilizados nestas combinações, auxiliam o pesquisador a traçar uma estratégia de busca mais efetiva.

Quadro 3 - Análise dos documentos recuperados nas bases de publicações científicas e patentárias segundo a metodologia de priorização *Facet Ranking*.

| Facetas | Número de Ocorrência | | |
|--------------|----------------------|----------------|-------------------|
| | Número de Facetas | Frequencia WoS | Frequencia Scopus |
| 0 | 6 | 3 | 24 |
| 1 | 80 | 33 | 26 |
| 2 | 246 | 149 | 279 |
| 3 | 137 | 194 | 121 |
| 4 | 28 | 93 | 44 |
| 5 | 2 | 26 | 4 |
| 6 | 1 | 2 | 2 |
| Total | 500 | 500 | 500 |

Fonte: Elaboração Própria.

As relações ou combinações das palavras-chave nos documentos recuperados estão representadas em uma matriz de relacionamento nos Quadros 4, 5 e 6 para as bases da Web of Science, Scopus e Derwent Innovation Index, respectivamente. Os valores iguais ou maiores a 30 estão destacadas em cor verde. Observa-se que as combinações mais frequentes na base do Web of Science foram entre as palavras *plant*, *seed*, *sorghum* e *sugarcane*, e com menor representatividade, próximo a 30, as palavras *germination* e bamb (bambu or bamboo). Não foi observado a recuperação da palavra *Saccharum* nesta base, indicando que o nome científico da cana-de-açúcar não é tão referenciado nestes documentos. A base da Scopus apresentou a maior quantidade de combinações, principalmente com os termos relacionados com a cana-de-açúcar, o que pode indicar que a fase de desenvolvimento tecnológico da propagação vegetativa da cana ainda está na fase científica. Os resultados na base do Derwent Innovation Index mostraram um comportamento mais semelhante a base do Web of Science, apesar da alta representatividade da palavra bamb. Estes resultados mostram que bases com o mesmo tipo de indexação, apresentam resultados similares, pois ambas WoS e DII pertencem a Thomson Reuters. Cabe destacar o maior volume de documentos de patente relacionados a propagação vegetativa de bambu em comparação aos artigos científicos.

Quadro 4 - Matriz de relacionamento entre as palavras-chave utilizadas na estratégia de busca para literatura científica e patentária sobre propagação vegetativa de cana-de-açúcar, sorgo e bambu na base do Web of Science. N = 500.

| WoS | Plant | Seed | Germination | Vegetative | Propagation | Plantlets | Sorghum | Sugarcane | Bamb | Saccharum |
|-------------|-------|------|-------------|------------|-------------|-----------|---------|-----------|------|-----------|
| Plant | 0 | 112 | 30 | 14 | 10 | 11 | 141 | 146 | 31 | 23 |
| Seed | 112 | 0 | 33 | 6 | 4 | 1 | 94 | 45 | 18 | 3 |
| Germination | 30 | 33 | 0 | 4 | 1 | 1 | 27 | 7 | 5 | 1 |
| Vegetative | 14 | 6 | 4 | 0 | 3 | 0 | 7 | 6 | 3 | 2 |
| Propagation | 10 | 4 | 1 | 3 | 0 | 2 | 3 | 6 | 6 | 1 |
| Plantlets | 11 | 1 | 1 | 0 | 2 | 0 | 0 | 8 | 1 | 1 |
| Sorghum | 141 | 94 | 27 | 7 | 3 | 0 | 0 | 11 | 1 | 3 |
| Sugarcane | 146 | 45 | 7 | 6 | 6 | 8 | 11 | 0 | 1 | 25 |
| Bamb | 31 | 18 | 5 | 3 | 6 | 1 | 1 | 1 | 0 | 1 |
| Saccharum | 23 | 3 | 1 | 2 | 1 | 1 | 3 | 25 | 1 | 0 |

Fonte: Elaboração própria.

Quadro 5 - Matriz de relacionamento entre as palavras-chave utilizadas na estratégia de busca para literatura científica e patentária sobre propagação vegetativa de cana-de-açúcar, sorgo e bambu na base do Scopus N=500.

Domjan, C.A.H; Winter, E.; Rodrigues, R.C. Estudo sistematizado para seleção de termos de busca para propagação vegetativa de cana-de-açúcar.

| Scopus | Plant | Seed | Germination | Vegetative | Propagation | Plantlets | Sorghum | Sugarcane | Bamb | Saccharum |
|-------------|-------|------|-------------|------------|-------------|-----------|---------|-----------|------|-----------|
| Plant | 0 | 121 | 36 | 28 | 34 | 62 | 85 | 309 | 4 | 144 |
| Seed | 121 | 0 | 36 | 12 | 11 | 19 | 56 | 108 | 3 | 40 |
| Germination | 36 | 36 | 0 | 3 | 5 | 5 | 19 | 32 | 1 | 21 |
| Vegetative | 28 | 12 | 3 | 0 | 6 | 2 | 12 | 14 | 1 | 2 |
| Propagation | 34 | 11 | 5 | 6 | 0 | 17 | 3 | 26 | 3 | 6 |
| Plantlets | 62 | 19 | 5 | 2 | 17 | 0 | 3 | 56 | 2 | 37 |
| Sorghum | 85 | 56 | 19 | 12 | 3 | 3 | 0 | 26 | 0 | 14 |
| Sugarcane | 309 | 108 | 32 | 14 | 26 | 56 | 26 | 0 | 0 | 165 |
| Bamb | 4 | 3 | 1 | 1 | 3 | 2 | 0 | 0 | 0 | 0 |
| Saccharum | 144 | 40 | 21 | 2 | 6 | 37 | 14 | 165 | 0 | 0 |

Fonte: Elaboração própria.

Quadro 6: Matriz de relacionamento entre as palavras-chave utilizadas na estratégia de busca para literatura científica e patentária sobre propagação vegetativa de cana-de-açúcar, sorgo e bambu na base do Derwent Innovation Index. N = 500.

| DII | Plant | Seed | Germination | Vegetative | Propagation | Plantlets | Sorghum | Sugarcane | Bamb | Saccharum |
|-------------|-------|------|-------------|------------|-------------|-----------|---------|-----------|------|-----------|
| Plant | 0 | 159 | 17 | 2 | 7 | 1 | 68 | 54 | 72 | 3 |
| Seed | 159 | 0 | 18 | 2 | 5 | 1 | 117 | 80 | 266 | 3 |
| Germination | 17 | 18 | 0 | 0 | 1 | 0 | 6 | 3 | 12 | 0 |
| Vegetative | 2 | 2 | 0 | 0 | 0 | 0 | 2 | 2 | 0 | 0 |
| Propagation | 7 | 5 | 1 | 0 | 0 | 0 | 2 | 4 | 1 | 0 |
| Plantlets | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| Sorghum | 68 | 117 | 6 | 2 | 2 | 0 | 0 | 35 | 15 | 3 |
| Sugarcane | 54 | 80 | 3 | 2 | 4 | 1 | 35 | 0 | 10 | 2 |
| Bamb | 72 | 266 | 12 | 0 | 1 | 0 | 15 | 10 | 0 | 2 |
| Saccharum | 3 | 3 | 0 | 0 | 0 | 0 | 3 | 2 | 2 | 0 |

Fonte: Elaboração própria.

Anterior a análise qualitativa, realizou-se uma avaliação da relevância dos termos nos documentos recuperados através de uma busca prévia, tanto em documentos patentários como não-patentários, nas bases do Web of Science e Derwent Innovation Index. Foram analisados 100 documentos de forma aleatória, onde foi possível constatar as seguintes observações:

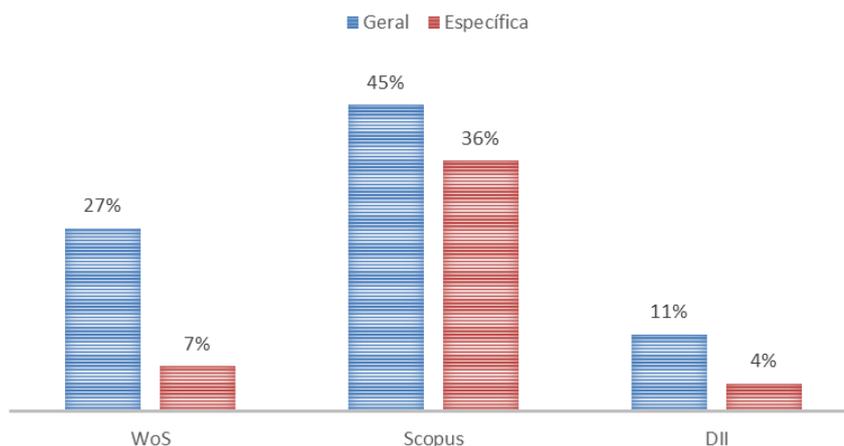
- 1) A palavra “*saccharum*” aparece associado tanto ao nome científico da cana-de-açúcar, sendo a mais comum a *Saccharum officinarum*, porém observaram-se outras formas de associação, tais como as palavras “*Acer*” e “*munja grass*”, as quais fogem muito do escopo da pesquisa;
- 2) O termo “*seedling*” foi mais específico para recuperar artigos relacionados com mudas do que o termo truncado *seed*;

Estas observações também ajudam na consolidação de uma estratégia de busca mais efetiva, recomendando a utilização de operadores booleanos NOT para os termos “*Acer*” e “*munja grass*”.

Em seguida, realizou-se a análise qualitativa em uma amostragem de 100 documentos a partir da priorização do *Facet Ranking*. Considerou-se dois parâmetros para a análise: a) relevância geral, definido como artigos ou patentes referenciando estudos sobre variedades de plantas ou uso de biocidas, os quais podem servir de subsidio para tecnologias de propagação vegetativa e b) relevância específico para metodologias de produção de mudas por propagação vegetativa de cana-de-açúcar, sorgo ou bambu. A Figura 2 mostra os resultados, no qual a base Scopus mostrou ser a mais precisa para recuperar documentos sobre a propagação vegetativa da cana-de-açúcar, com 36,5% dos documentos relevantes no universo analisado qualitativamente. Apesar do tamanho

amostral relativamente pequeno em proporção ao volume das informações recuperados, estes resultados mostram as diferenças de cobertura (Figura 1) e a precisão da recuperação das informações nas diferentes bases (Figura 2), corroborando que a fase de desenvolvimento tecnológico da propagação vegetativa da cana-de-açúcar ainda está na fase de pesquisa.

Figura 2: Porcentagem de documentos relevantes para a propagação vegetativa, geral e específico, segundo a base de dados consultada: Web of Science (WoS), Scopus e Derwent Innovation Index (DII).



Fonte: Elaboração própria.

CONCLUSÕES E PERSPECTIVAS

A principal contribuição deste trabalho é mostrar a importância em realizar estudos sistematizados para a escolha de termos destinados na recuperação de documentos científicos e de patente de forma mais eficiente, capazes de subsidiar a pesquisa e desenvolvimento na área agroindustrial, especificamente para tecnologias de propagação vegetativa. Ademais, o segmento estudado mostrou ser altamente relevante e promissor para o setor produtivo de cana-de-açúcar, refletido pelo alto número de artigos científicos e documentos de patente nesta área.

A metodologia adotada possibilitou identificar os termos ou palavras-chave mais precisas para recuperação de documentos de propagação vegetativa da cana-de-açúcar, destacando-se “*plant*” nas bases de artigos científicos e “*seed*” na base de patentes. Já as palavras “*vegetative*” e “*propagation*” não foram representativas em nenhuma base consultada, apesar de serem literalmente a tradução do termo “propagação vegetativa”. Os termos “*germination*” e “*plantlets*” mostraram ser representativas na base de artigos científico Scopus, ao igual que o termo “*Saccharum*”. O termo “*sugarcane*” mostrou ser mais preciso para recuperar documentos sobre cana-de-açúcar do que *Saccharum* nas demais bases. Muitas destas diferenças são relacionadas a indexação das bases e diferenças no linguajar técnico das bases de patentes frente às bases de literatura científica.

Em relação a utilização de espécies que sirvam como *benchmark* nas pesquisas sobre propagação vegetativa de cana-de-açúcar, observou-se que o sorgo (*Sorghum*) é mais preciso em artigos científicos, enquanto que tanto o sorgo como o bambu podem ser úteis na literatura patentária.

Pode-se concluir que a base Scopus teve uma precisão melhor na recuperação de documentos para a tecnologia de propagação vegetativa de cana-de-açúcar. Entretanto, recomenda-se que o

levantamento de informações não seja limitado a uma única base, contemplando pesquisas em bases de patentes que possam contribuir com maior objetividade para o setor agroindustrial.

Como perspectivas futuras, recomenda-se avaliar se as diferenças nas bases de artigos e de patentes realmente reflete o estado de desenvolvimento da tecnologia de propagação vegetativa, bem como verificar se a cobertura maior da Scopus está relacionada com um maior interesse dos países latinos na produção de cana-de-açúcar.

AGRADECIMENTOS

Bolsa CNPq – Convênio INPI-CNPq.

REFERÊNCIAS

ALBERTS D. et al. Dominic. Introduction to Patent Searching: Practical Experience and Requirements for Searching the Patent Space. **The Information Retrieval Series**, V.29, pp 3-43, 2011.

ALBUQUERQUE, Eduardo da Motta; SIMÕES, Rodrigo; BAESA, Adriano; CAMPOLINA, Bernardo; SILVA, Leandro. A Distribuição Espacial da Produção Científica Tecnológica Brasileira: uma Descrição de Estatísticas de Produção Local de Patentes e Artigos Científicos. **Revista Brasileira de Inovação**, v. 1, n. 2 jul/dez, p. 225-251,2002.

BONINO, Dario; CIARAMELLA, Alberto; CORNO, Fulvio. Review of the state-of-the-art in patent information and forthcoming evolutions in intelligent patent informatics. **World Patent Information**, v.32, p.30-38, 2010.

CAO, Yang; FAN, Ju; LI, Guoliang. A User-Friendly Patent Search Paradigm. **IEEE Transactions On Knowledge and Data Engineering**, V. 25, N.6, pg. 1439-1443, jun, 2013.

CARVALHO, S. A. D.; FURTADO, A. T. Mapeamento dos esforços tecnológicos dos programas de melhoramento genético de cana de açúcar no Brasil. In: ALTEC 2013 – XV Congresso da Associação Latino-Iberoamericana de Gestão de Tecnologia, 2013, Porto. **Anais do Congresso da ALTEC 2013**, 2013.

CTC – CENTRO DE TECNOLOGIA CANAVIEIRA. Método para obtenção de propágulos vegetais para micropropagação de cana-de-açúcar, para produção de semente sintética de cana-de-açúcar, para armazenamento de semente sintética, para produção de mudas, propágulos viáveis, e, semente sintética de cana-de-açúcar. **BR 10 2013 007727 5 A2**, 68 p.,14 jul. 2015.

DIAS, Cleber Gustavo, ALMEIDA, Roberto Barbosa de. Produção científica e produção tecnológica: transformando um trabalho científico em pedidos de patente. **Einstein** (São Paulo) [online], vol.11, n.1, pp. 1-10, 2013. Disponível em: http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1679-45082013000100003. Acesso: outubro de 2015.

FAO. Estatística da produção mundial de cultivo de cana-de-açúcar, ano base 2014. Disponível em: <http://faostat3.fao.org/browse/Q/QC/E> Acesso: 20 mai 2016.

Domjan, C.A.H; Winter, E.; Rodrigues, R.C. Estudo sistematizado para seleção de termos de busca para propagação vegetativa de cana-de-açúcar.

FERREIRA, Ademir Antônio; GUIMARAES, Edílson Rodrigues and CONTADOR, José Celso. Patente como instrumento competitivo e como fonte de informação tecnológica. **Gest. Prod.** [online], vol.16, n.2, pp. 209-22, 2009. Disponível em: http://www.scielo.br/scielo.php?pid=S0104-530X2009000200005&script=sci_abstract&tlng=pt Acesso: outubro de 2015.

GOEDKEN J. F. Methods and apparatus to search and analyze prior art. **US2005210042 (A1)**, 44p., 22 set. 2005.

GUO, Yunsong; GOMES, Carla P. Ranking Structured Documents: A Large Margin Based Approach for Patent Prior Art Search. **IJCAI, USA**, jul., pg.11-17, 2009.

HUNT, David; NGUYEN, Long; RODGERS, Matthew. Patent Searchig: Tools & Techniques. **John Wiley&Sons, Inc.** p. 188, 2007.

KOBE UNIV, & KANEKA CORP. Patent Ranking Device and Patent Ranking Method. **JP 2014199661 (A)**, 39p, 23 out. 2014.

LANDELL *et al.* Sistema de multiplicação de cana-de-açúcar com uso de muda pré-brotadas (MPB) oriundas de gemas individualizadas. **Documentos IAC**, Campinas, 109, 2012.

MAGDY, Walid; JONES, Gareth J. F. Examining the Robustness of Evaluation Metrics for Patent Retrieval with Incomplete Relevance Judgements, **CLEF**, 2010.

MARICATO, João de Melo. NORONHA, Daisy Pires. Coclasseificação em artigos e patentes em biodiesel. **Linc em Revista**, Rio de Janeiro, v. 9, n. 1, p. 85-102, maio 2013. Disponível em: <http://revista.ibict.br/liinc/index.php/liinc/article/view/510/391> Acesso: outubro de 2015.

MATERNE, Alain; SLEIGHTHOLME, Gershom. Methos of ranking search results for searchers based on multiple search concepts carried out in multiple database. **World Patent Information**, v. 36, pg. 4-15, 2014.

MOELLER, Ansgar; MOEHRLE, Martin G. Completing keyword patent search with semantic patent search: introducing a semiautomatic iterative method for patent near search based on semantic similarities. **Scientometrics**, V.102, pg.77-96, 2015.

MONCHAUX, Sophie; AMADIEU, Franck; CHEVALIER, Aline; MARINÉ, Claudette. Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. **Information Processing and Management**, V.51, pg.557-569, 2015.

MOURA, Ana Maria Mielniczuk de. **A interação entre artigos e patentes: um estudo cientométrico da comunicação científica e tecnológica em biotecnologia**. Tese de Doutorado. Universidade Federal do Rio Grande do Sul. Faculdade de Biblioteconomia e Comunicação. Programa de Pós-Graduação em Comunicação e Informação, 269 f. 2009.

NEVES, M. F.; TROMBIN, V. G.; SIMPRINI, E. S.; RODRIGUES, O. M.; CANTO, F.; ROVANHOL, P.; CONSOL, M. H. A dimensão do setor sucroenergético: mapeamento e quantificação da safra 2013/14. 1. Ribeirão Preto: **Markestrat, Fundace, FEA-RP/USP**, v. 1. 46p, 2014.

Domjan, C.A.H; Winter, E.; Rodrigues, R.C. Estudo sistematizado para seleção de termos de busca para propagação vegetativa de cana-de-açúcar.

NIEVES N., ZAMBRANO Y., TAPIA R., CID M., PINA D., CASTILLO R. Field performance of artificial seed-derived sugarcane plants. **Plant Cell, Tissue and Organ Culture**, v.75, n.3, p. 279-282, 2003.

NIJHOF, Evert. Searching? Or trying to find something? The comforts of searching versus the challenges of finding. **World Patent Information**, v. 33, p. 360-363, 2011.

OLIVEIRA, Luciana Goulart de et al. Informação de patentes: ferramenta indispensável para a pesquisa e o desenvolvimento tecnológico. **Quím. Nova**, São Paulo, v. 28, supl. p. S36-40, dez. 2005.

RAVASCHIO, Juliana de Paula; FARIA, Leandro Innocentini Lopes de; QUONIAM, Luc. O uso de patentes como fonte de informação em dissertações e teses de engenharia química: o caso da Unicamp. **Rev. digit. bibliotecon. cienc. inf.** v. 7, n. 2, p. 219-232, 2010.

ROBERTSON, S. E. The probability ranking principle in IR. *J. Documentation*, v. 33, n.4, pp. 294-304, 1977.

SALTON, Gerard; BUCKLEY, Christopher. Term-weighting approaches in automatic text retrieval. **Information Processing and Management**, p.513-523, 1988. Disponível em: <http://www.cs.odu.edu/~jbollen/IR04/readings/article1-29-03.pdf> Acesso: outubro, 2015.

SILVEIRA, Murilo Artur Araújo da. **O artigo científico como fonte de informação utilizada na literatura cinzenta**. Monografia, Universidade Federal de Pernambuco. Recife, 2005.

VALOR. Novas tecnologias buscam modernizar o plantio de cana. Publicado em: 26 ago 2014. Disponível em: <http://alfonsin.com.br/novas-tecnologias-buscam-modernizar-o-plantio-de-cana/> Acesso: 02/02/2015

WINTER, E.; LIMA, Araken A.; MENDES, C. D. S. Mapeamento tecnológico da cadeia produtiva do etanol proveniente da cana-de-açúcar sob enfoque dos pedidos de patentes: cenário brasileiro. In: Luís Augusto Barbosa Cortez. (Org.). *Bioetanol de Cana-de-Açúcar: P&D para produtividade e sustentabilidade*. São Paulo: **Editora Edgar Blucher**; 2010.

ZMC INTELLECTUAL DATA SCIENCE & TECHNOLOGY CO. LTD. Ranking analysis system and method of patent family numbers of applicants. **CN 101996177 (A)**, 6p., 30 mar. 2011.