

PATENT ANALYSIS: AN APPROACH USING BIBLIOMETRIX.

**ANÁLISE DE PATENTES: UMA ABORDAGEM METODOLÓGICA USANDO O
BIBLIOMETRIX**

Eduardo Amadeu Dutra Moresi

Universidade Católica de Brasília

Isabel Pinho ; <https://orcid.org/0000-0003-1714-8979>

Universidade de Aveiro



Patent analysis: an approach using Bibliometrix Análise de patentes: uma abordagem metodológica usando o Bibliometrix

Abstract: Patent application documents are systematically collected in a structured data format (including application date, applicant and inventor name, and a complete technical description of the invention) in publicly available patent databases. The purpose of this paper is to present a guide that exploits the bibliometric analysis capabilities available in the R-Bibliometrix package for patent analysis. This objective will broaden the understanding of quantitative results and allow further interpretation of research findings from selected documents by means of bibliometric metrics. Therefore, the purpose of this work is to answer the following research question: how to employ bibliometric analysis in the interpretation of documents retrieved in searches carried out in patent databases? A search was conducted in the Lens database, which is open access, on the topic of virtual worlds. The application of the approach allowed us to identify some performance indicators and to trace a technological map, including the annual evolution of patent production; the quantity of documents per jurisdiction; the main patent owners; the main inventors and their institutions; the inventors' production in the period 2006 to 2022; the most frequent keywords extracted from the abstracts; the most frequent CPC class codes; the technological trends from the CPC subclasses; the thematic maps, and the evolution of the technological structure of the CPC subclasses and the bigrams extracted from the abstracts. It's concluded by highlighting the potential of the R-Bibliometrix package for patent analysis to identify performance indicators and perform technological mapping.

Keywords: patent analysis, bibliometrics, Bibliometrix, performance indicators, technology mapping.

Resumo: Os documentos de pedido de patente são sistematicamente coletados em um formato de dados estruturados (incluindo data do pedido, nome do depositante e do inventor, e uma descrição técnica completa da invenção) em bancos de dados de patentes disponíveis ao público. O propósito deste trabalho é apresentar um guia que explore as potencialidades de análise bibliométrica disponíveis no pacote R-Bibliometrix para a análise de patentes. Este objetivo irá ampliar o entendimento dos resultados quantitativos e permitir aprofundar a interpretação dos resultados da pesquisa a partir de documentos selecionados por meio de métricas bibliométricas. Portanto, a proposta deste trabalho é responder a seguinte questão de pesquisa: como empregar a análise bibliométrica na interpretação de documentos recuperados em pesquisas realizadas em bases de patentes? Foi realizada uma pesquisa na base Lens, que é de acesso livre, sobre o tema mundos virtuais. A aplicação da abordagem permitiu identificar alguns indicadores de desempenho e traçar um panorama tecnológico do tema, incluindo a evolução anual da produção de patentes, os quantitativos de documentos por jurisdição, os principais detentores de patentes, os principais inventores com os respectivos vínculos institucionais, a produção dos inventores no período de 2006 a 2022, as palavras-chave mais frequentes extraídas dos resumos, os códigos das classes CPC mais frequentes, as tendências tecnológicas a partir das subclasses CPC, os mapas temáticos e as evoluções da estrutura tecnológica das subclasses CPC e dos bigramas extraídos dos resumos. Conclui-se ressaltando as potencialidades do pacote R-Bibliometrix para a análise de patentes visando identificar indicadores de desempenho e realizar o mapeamento tecnológico.

Palavras-chave: análise de patentes, bibliometria, Bibliometrix, indicadores de desempenho, mapeamento tecnológico.

1. Introdução

As patentes são úteis para análise competitiva e de tendências tecnológicas (Abraham & Morita, 2001), tendo sido empregadas na gestão de projetos de pesquisa e desenvolvimento para avaliar a posição competitiva da indústria. Os pedidos de patentes estão disponíveis ao público e incluem informações úteis, tais como o nome do requerente, data da aplicação e detalhes técnicos da invenção. Trata-se de uma fonte de informação valiosa para auxiliar a avaliação de tendências na pesquisa, uma vez que revelam as áreas de inovação que os inventores estão focados. A análise de patentes também é uma abordagem interessante para estudos prospectivos por utilizar informações textuais sobre propriedade intelectual de diversas áreas do conhecimento.

Os dados de patentes são muito valiosos ao rastrear tendências de pesquisa e inovação como fonte de confiáveis de informações técnicas, jurídicas e comerciais. Os documentos de pedido de patente são sistematicamente coletados em um formato de dados estruturados (incluindo data do pedido, nome do depositante e do inventor, e uma descrição técnica completa da invenção) em bancos de dados de patentes disponíveis ao público. Portanto, eles são uma boa fonte de dados para análise, da qual podem ser extraídas conclusões significativas.

A análise de dados de patentes permite acompanhar as mudanças ao longo do tempo e identificar onde a maioria das patentes estão sendo depositadas e por quem. Também são uma rica fonte de informações técnicas. Investigar mais a fundo a substância das aplicações permite ter uma visão dos tipos de tecnologias que estão sendo desenvolvidas e aquelas que estão surgindo, como elas são aplicadas e os campos que cobrem. Isto ajuda a compreender melhor o perfil atual e a direção futura de uma indústria.

Os documentos de patentes contêm informações técnicas sobre desenvolvimentos em um determinado campo que muitas vezes é único. Isto se deve ao fato de que os requerentes de patentes - em primeiro lugar e principalmente as empresas - muitas vezes buscarão proteção de patentes sem publicação de artigos científicos relacionados. Os dados de patentes também contêm inerentemente informações comerciais relevantes. Um solicitante não apenas investe na pesquisa e desenvolvimento de sua invenção, mas arca com o custo do depósito da patente em cada escritório jurisdicional (já que a proteção de patente é territorial). Isso significa que só vale a pena fazer naquelas jurisdições onde existe um mercado potencial para soluções protegidas.

O crescimento de patentes geralmente segue uma tendência semelhante que pode se assemelhar ao crescimento em forma de S (Chen, Chen & Lee, 2010). Nos estágios iniciais de uma tecnologia, o número de patentes emitidas é muito limitado. Segue-se um período de rápido crescimento quando o número de patentes depositadas e submetidas aumenta e então é atingido um patamar de maturidade. Como o processo de registro de patentes é caro e pode levar vários anos, o depósito em um escritório de registro de propriedade intelectual significa que há otimismo na contribuição econômica ou técnica do conhecimento que será protegido. Através da análise das informações registradas em uma patente, é possível acompanhar as mudanças ao longo do tempo e identificar quais as jurisdições, ou seja, os escritórios de propriedade intelectual dos países, estão realizando a maioria das atividades de patenteamento. Além disso, aprofundando a análise do conteúdo em que é requerida a

proteção em termos de patente, é possível ter uma visão dos tipos de tecnologias em desenvolvimento e aquelas que estão surgindo e onde serão aplicadas.

Vários índices foram introduzidos para medir a força tecnológica em função da quantidade ou qualidade da patente. Alguns exemplos incluem índices de citação de patentes e modelos de regressão (Wantanabe, Tsuji & Brown, 2001). A análise de patentes tem se mostrado valiosa no planejamento do desenvolvimento tecnológico, desde o estudo da estratégia em nível nacional (Abraham & Morita, 2001) até a modelagem de tecnologias emergentes específicas (Bengisu & Nekhili, 2006). Os metadados de patentes são geralmente de acesso livre na maioria dos países e várias diretrizes foram introduzidas para aprimorar a técnica usando palavras-chave e categorização.

Todavia, os registros nas bases de patentes não fornecem as palavras-chave, sendo que a categorização pode ser feita utilizando os códigos de classificação de patentes. Há duas classificações de patentes usadas em âmbito internacional: a Classificação Internacional de Patentes (IPC – *The International Patent Classification*), disponível no sítio da World Intellectual Property Organization (2022) e a Classificação de Cooperação de Patentes (CPC – *The Cooperative Patent Classification*), que pode ser consultada no European Patent Office (2022). A classificação de um documento de patente influencia todo o ciclo de vida deste documento. A cada nova versão da IPC e da CPC, novos símbolos são criados, mostrando a tendência do desenvolvimento tecnológico nas áreas afins.

Em suma, os bancos de patentes podem ser utilizados para: sondar as atuais tendências por meio do conhecimento do estado da técnica e histórico da tecnologia; visualizar os inventos mais recentes aplicados nas mais diversas áreas do desenvolvimento de pesquisas; prospectar uma determinada tecnologia; levantar novos temas para trabalhos acadêmicos; identificar o nível de exploração de uma tecnologia; buscar novas maneiras técnicas para a solução de problemáticas existentes na sua área de atuação; dentre outras utilizações possíveis.

A bibliometria é definida por Norton (2010) como a medida de textos e informações, tornando-se uma ferramenta padrão para a gestão da pesquisa nas últimas décadas. Todas as compilações significativas de indicadores científicos dependem fortemente da publicação e estatísticas de citação e outras técnicas bibliométricas mais sofisticadas. Os métodos bibliométricos têm sido usados para rastrear as citações de periódicos acadêmicos. Entretanto, hoje a bibliometria pode ser usada para entender o passado e até mesmo potencialmente para estimar o futuro (Morris et al, 2022). A bibliometria ajuda a explorar, organizar e analisar grandes quantidades de dados históricos, ajudando os pesquisadores a identificar padrões ocultos que podem ajudar os pesquisadores no processo de tomada de decisão. Algumas técnicas analíticas têm sido usualmente utilizadas em bibliometria, tais como autores, afiliações, mapas conceituais, análise de agrupamento e fator, análise de citação e cocitação para mencionar algumas delas.

Na análise de patentes, existem diversos trabalhos que empregaram a análise bibliométrica. Entre os diversos exemplos podemos mencionar: a construção de um modelo de análise de patentes para orientar a transferência de tecnologia (Liu & Yu, 2017); o desenvolvimento de um modelo de análise de patentes para promover o desdobramento da tecnologia de grafeno na indústria de TIC (Xi & Xiang, 2020); a indicação de previsões para três áreas tecnológicas emergentes integrando a análise de patentes com ferramentas de prospecção

(Daim et al, 2006); o mapeamento tecnológico utilizando análise de patentes e de redes sociais para identificar especialistas e tecnologias existentes ou em desenvolvimento (Daim et al, 2014); em uma revisão de literatura foi verificado o potencial industrial dos biocatalisadores, particularmente das lipases (Almeida et al, 2021); a introdução e a previsão de biossensores para uso em aplicações da área de saúde com o emprego de biossensores (Sheikh & Sheikh, 2017), uma meta-análise sobre o desenvolvimento de pesquisas sobre análise de patentes no período de 1980 a 2003 (Lai et al, 2007); o fornecimento de uma abordagem sistemática para a análise integrada de tendências que levam em conta a análise de patentes empregando mineração de textos (Wu et al, 2011).

Contudo, não foram identificados estudos utilizando o pacote R-Bibliometrix nas análises quantitativa e qualitativa de informações registradas em patentes. O propósito deste trabalho é apresentar um guia que explore as potencialidades de análise bibliométrica disponíveis no pacote R-Bibliometrix para a análise de patentes. Este objetivo irá ampliar o entendimento dos resultados quantitativos e permitir aprofundar a interpretação dos resultados da pesquisa a partir de documentos selecionados por meio de métricas bibliométricas. Portanto, a proposta deste trabalho é responder a seguinte questão de pesquisa: como empregar a análise bibliométrica na interpretação de documentos recuperados em pesquisas realizadas em bases de patentes?

2. Metodologia

Este trabalho propõe uma abordagem para explorar as informações recuperadas em bases de patentes. A Figura 1 apresenta uma abordagem para explorar as patentes, incluindo as seguintes etapas: desenho da pesquisa; coleta e compilação dos dados de patentes; preparação e importação dos metadados; análise bibliométrica, incluindo a análise de desempenho e o mapeamento tecnológico; análise qualitativa e relato dos resultados da pesquisa. Esta abordagem permite explorar os indicadores disponíveis nas bases Derwent Innovations Index e Lens, além daqueles que podem ser obtidos pelo Bibliometrix, tais como: dinâmica de documentos por jurisdição, número de documentos por inventores ou proprietários em um período de tempo. O mapeamento tecnológico pode ser explorado mediante análise da estrutura tecnológica ou do ciclo de desenvolvimento tecnológico. Por fim, pode ser utilizada a análise qualitativa para fundamentar uma análise crítica do panorama tecnológico.

A primeira etapa em qualquer estudo bibliométrico é o desenho da pesquisa. É preciso definir a questão norteadora da pesquisa e escolher um método bibliométrico apropriado para respondê-la. Diferentes métodos são adequados para responder a diferentes questões de pesquisa. Algumas questões de pesquisa são apresentadas no Quadro 1. As perguntas se basearam nas análises de desempenho e no mapeamento tecnológico.

Considerando a necessidade de exportação dos metadados, sugere-se que a pesquisa seja realizada nas bases Derwent Innovations Index (DII) ou Lens. A base DII fornece acesso a mais de 30 milhões de invenções detalhadas em mais de 65 milhões de documentos de patentes, indexadas de 59 escritórios internacionais, além de incluir os links para patentes citadas, artigos citados e indexados na base Web of Science e texto integral das patentes (Clarivate, 2021). O acesso à base DII pode ser feito via Portal CAPES.

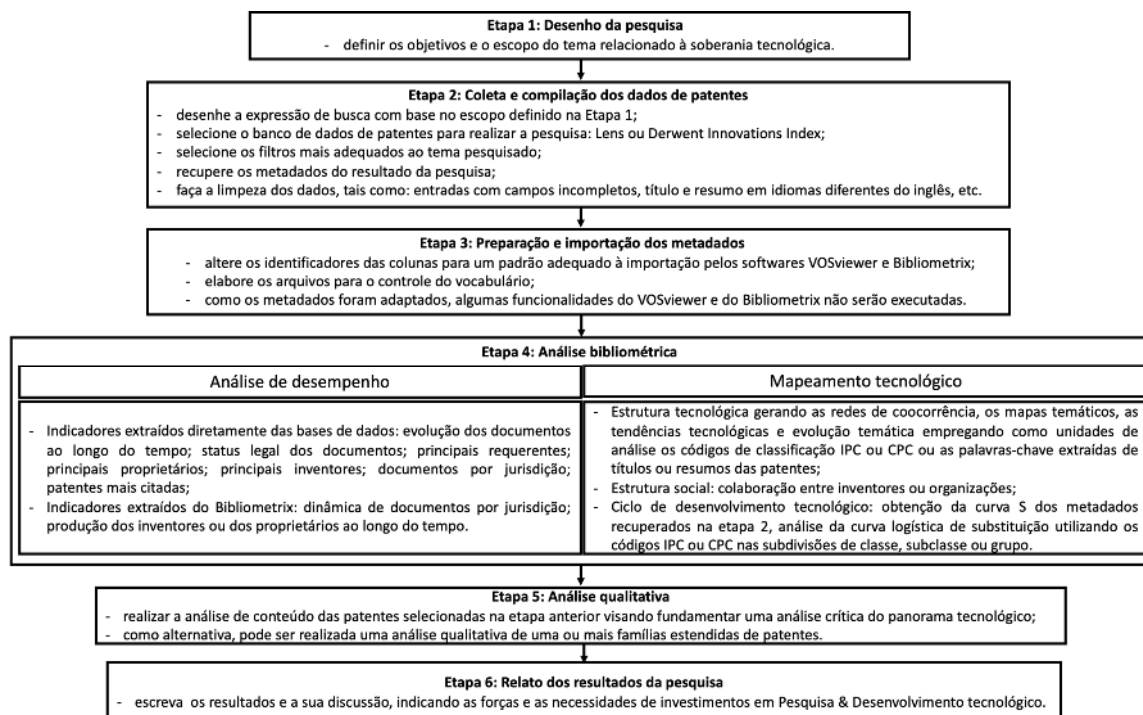


Figura 1 – Abordagem para a pesquisa e a análise de panorama tecnológico.

Outra opção é a pesquisa de patentes utilizando a base Lens (2022), que é de acesso livre e indexa documentos científicos e patentes. A Lens tem uma atualização semanal da indexação de patentes. A Lens é um agregador de metadados, combinando três conjuntos de conteúdo único e uma ferramenta de gerenciamento como uma oferta de base. Esta base suporta as quatro principais funções voltadas para a pesquisa, que são descobrir, analisar, gerenciar e compartilhar conhecimentos. Considerando os recursos disponíveis e o acesso gratuito, a base Lens foi escolhida para a pesquisa e a exportação dos metadados para a análise de documentos de patentes.

Quadro 1 – Questões norteadoras de pesquisa respondidas por diferentes métodos para análise de patentes.

Método	Questões norteadoras
Indicadores de Desempenho	Como estão as tendências de registro de patentes no período abrangido pela pesquisa? Quais são as principais indústrias envolvidas nesta tecnologia? Quem são os principais agentes de mercado envolvidos e seus países designados? Quais são as tendências emergentes no âmbito temas pesquisado? Quais são as principais jurisdições de registro de patentes?
Estrutura social	<ul style="list-style-type: none"> - Quais os inventores mais centrais ou de intermediação na colaboração na tecnologia pesquisada? - Quais as indústrias mais centrais ou de intermediação que colaboram em um campo de pesquisa? - Qual é a estrutura social do campo?
Estrutura tecnológica	<ul style="list-style-type: none"> - Quais são as dinâmicas da estrutura conceitual de um campo? - Quais são os tópicos associados a uma determinada tecnologia? - Qual a evolução de um determinado tópico?

Fonte: elaborado pelos autores.

Uma variedade de medidas de relação entre documentos, autores, organizações e palavras foi proposta e usada na literatura (Klavans & Boyack, 2006) para a interpretação de artigos e podem ser usados para as patentes. Para realizar essas medidas de relações, vários softwares podem ser utilizados. Neste trabalho, são sugeridas duas possibilidades:

- o pacote R Bibliometrix;
- os softwares VOSviewer para gerar as redes (van Eck & Waltman, 2021) e o Gephi para calcular as métricas de redes (Bastian, Heymann, & Jacomy, 2009).

Neste trabalho será o usado o pacote R Bibliometrix (Aria & Cuccurullo, 2021), que possui diversas funcionalidades para a análise de desempenho e o mapeamento tecnológico. Este aplicativo também oferece outros recursos para mapas de agrupamentos de documentos e estruturas tecnológica e social. Cabe ressaltar que, como os metadados devem ser ajustados para o padrão de artigos. Como as bases não exportam as referências bibliográficas, algumas funcionalidades não funcionarão e não será possível analisar a estrutura intelectual com o pacote R-Bibliometrix.

A estrutura tecnológica se baseia na coocorrência de palavras e oferece as seguintes opções:

- rede de coocorrência de palavras-chave dos autores, palavras-chave plus, título (monograma, bigrama e trigrama) e resumo (monograma, bigrama e trigrama), com a visualização dos mapas, das tabelas de dados das redes e dos gráficos de grau;
- mapa temático de palavras-chave dos autores, palavras-chave plus, título (monograma, bigrama e trigrama) e resumo (monograma, bigrama e trigrama), com a visualização dos mapas, das redes e das tabelas de dados e dos agrupamentos;
- evolução temática de palavras-chave dos autores, palavras-chave plus, título (monograma, bigrama e trigrama) e resumo (monograma, bigrama e trigrama), com a visualização dos mapas temáticos, das redes e das tabelas de dados e dos agrupamentos¹;
- análise fatorial que utiliza os métodos de análise de correspondência, de análise de correspondência múltipla e de dimensionamento multidimensional para palavras-chave dos autores, palavras-chave plus, título (monograma, bigrama e trigrama) e resumo (monograma, bigrama e trigrama), com a visualização do mapa de palavras, do dendrograma de tópicos, dos artigos que mais contribuíram, dos artigos mais citados, dos agrupamentos por palavras, dos artigos por agrupamentos.

A estrutura social oferece as seguintes opções: rede colaboração de autores, instituições e países, com a visualização dos grafos, das tabelas de dados e dos gráficos de grau; e mapa de colaboração de países, com a visualização do grafo e da tabela de dados.

Todavia, os registros nas bases de patentes não fornecem as palavras-chave, sendo que a categorização pode ser feita utilizando os códigos de classificação de patentes. Há duas classificações de patentes usadas em âmbito internacional: a Classificação Internacional de Patentes (IPC – The International Patent Classification), disponível no sítio da World Intellectual Property Organization (2022) e a Classificação de Cooperação de Patentes (CPC – The Cooperative Patent Classification), que pode ser consultada no European Patent Office (2022). A classificação de um documento de patente influencia todo o ciclo de vida deste documento. A cada nova versão da IPC e da CPC, novos símbolos são criados, mostrando a tendência do desenvolvimento tecnológico nas áreas afins.

Estes códigos e os seguintes símbolos combinados: a seção; a classe; a subclasse; o grupo principal e o subgrupo. As seções são o nível mais alto da hierarquia da classificação e são divididas em oito. Cada uma delas é simbolizada por uma letra maiúscula de A até H, seguidas pelo título da Seção, que é uma indicação ampla do seu conteúdo: A - Necessidades

¹ A evolução temática permite a visualização por períodos de tempo.

Humanas; B - Operações de Processamento, Transporte; C – Química, Metalurgia; D – Têxteis, Papel; E - Construções Fixas; F - Engenharia Mecânica, Iluminação, Aquecimento, Armas, Explosão; G – Física; e, H – Eletricidade. O CPC possui uma seção adicional denominada Y - Tecnologias transversais emergentes.

O segundo nível hierárquico são as chamadas Classes, que possuem dois dígitos numéricos. O terceiro nível são as Subclasses, que são codificadas por um dígito alfabético, cujo título indica de forma um pouco mais precisa o seu conteúdo. No quarto nível estão os Grupos, que se subdividem entre Grupos principais e Subgrupos. Os Grupos Principais são representados por até dois dígitos numéricos seguido de uma barra oblíqua e o número 00. Além disso, cada Grupo Principal possui um título que define o conteúdo dentro do escopo da sua subclasse para facilitar a pesquisa de documentos com assuntos mais específicos. Em algumas Subclasses existe o Grupo 99/00, que inclui matéria não abrangida pelos demais Grupos de uma Subclasse. Os Subgrupos são as subdivisões dos Grupos Principais e são representados por pelo menos dois dígitos numéricos diferentes de 00. Por exemplo, o código G06N 20/20 pode ser interpretado como: G – Física; G06 – Cômputo; Cálculo ou Contagem; G06N – Disposições de computação baseadas em modelos computacionais específicos; G06N 20/00 – Aprendizado de Máquina; G06N 20/20 – Aprendizagem de conjunto.

Os códigos IPC e CPC são incluídos até o nível de subgrupo. Contudo, a análise do panorama tecnológico poderá exigir obter redes ou mapas nos níveis de classe, subclasse ou grupo. Para isso será aplicada uma conversão dos códigos IPC ou CPC das patentes para níveis imediatamente superiores.

O outro procedimento envolve os softwares VOSviewer e Gephi. O VOSviewer é um software livre especialmente projetado para gerar representação gráfica de mapas bibliométricos (van Eck & Waltman, 2014; Waltman & van Eck, 2012). Ele inclui funcionalidades de zoom, de algoritmos especiais de tabulação e de mapa de densidade. O software pode ser usado para construir e visualizar mapas bibliométricos de dados de coautoria (autores, organizações e países) e coocorrência de palavras-chave (dos autores e plus). Além disso, o VOSviewer permite realizar a detecção de comunidades usando a técnica de agrupamento. O VOSviewer permite seu exame através de 3 tipos de visualização: de rede, de sobreposição e de densidade (van Eck & Waltman, 2021).

O Gephi, como software de análise de redes complexas, é frequentemente utilizado para calcular as métricas de redes. Com ele é possível calcular o grau médio, que determina o grau de todos os nós, e as centralidades de proximidade, intermediação e autovetor (Bastian, Heymann, & Jacomy, 2009). Outros parâmetros também poderão ser calculados, tais como: classes de modularidade, excentricidade, número de triângulos, etc. O Gephi permite a visualização do grafo da rede com a atribuição de cores para cada agrupamento ou classe de modularidade e algoritmos de layout (Kauffman et al., 2014): Force Atlas, Force Atlas 2, Fruchterman Reingold, OpenOrd, Yifan Hu e Yifan Hu Proporcional. O Gephi permite visualizar e manipular os dados por meio da funcionalidade de Laboratório de Dados. Várias ações são possíveis, tais como ordenar os nós pelas métricas calculadas, editar os nós, incluir ou excluir colunas, visualizar arestas, etc.

3. A base Lens

A base Lens tem o objetivo de conectar artefatos de conhecimento aberto e metadados com ferramentas para auxiliar a solução eficaz, eficiente e equitativa de problemas. A base abrange:

- obras acadêmicas: ferramentas de descoberta e análise que fornecem acesso a um corpus global de metadados da literatura acadêmica com indexação de citações;
- Patentes: ferramentas de descoberta e análise de uma coleção abrangente de literatura de patentes com indexação de citações;
- PatSeq: uma facilidade para pesquisar e analisar sequências biológicas divulgadas na literatura de patentes;
- Coleções: uma ferramenta de gerenciamento para rastrear, monitorar e analisar uma coleção de obras ou uma coleção de patentes de forma dinâmica ou estática.

Estas ferramentas fundamentais têm acesso universal, ou seja, estão disponíveis para usuários que se registram e para aqueles que não se registram. Elas são abertas, livres e permitem o acesso privado e seguro a diferentes fontes de informação. Para ter acesso a mais recursos, como salvar pesquisas e exportar os metadados em diversos formatos, limitados a 50 mil registros de cada vez.

Atualmente, a Lens abrange 68.347.718 documentos de patentes concedidas e 64.735.922 de submetidas; mais de 100 escritórios de registro de propriedade intelectual, incluindo China, Estados Unidos, Japão, Alemanha, Escritório Europeu (European Patent Office), Escritório da Organização Mundial de Propriedade Intelectual (World Intellectual Property Organization – WIPO), além, de Brasil, Portugal e Espanha.

A base Lens também indexa mais de 247 milhões de registros acadêmicos, compilados e harmonizados das bases Microsoft Academic, PubMed, OpenAlex e Core. Os documentos são enriquecidos com informações de acesso aberto via Unpaywall (Piwowar et al, 2018), texto completo CORE (2022) e links de identificação de autores pelo ORCID (2022). O acesso livre e o conceito central da base Lens, focado em ciência aberta, foram considerados como fundamentais, para que pesquisadores, sem acesso às bases pagas com informações sobre patentes e artigos científicos possam ter acesso a estas informações e conduzir suas pesquisas.

É importante ressaltar que a base Lens possui os seguintes recursos para apoiar a pesquisa e a interpretação dos resultados:

- filtros: intervalo de tempo; identificadores de metadados; escritórios de jurisdição; aplicantes; inventores; proprietário da patente; status legal da patente; tipo de documento (patente concedida, pedido de patente, patente limitada, etc); trabalhos citados; códigos de classificação; família de documentos em uma jurisdição; ferramentas de consultas;
- ferramentas de análise: documentos de patentes ao longo do tempo; documentos de patentes por data de publicação, depósito e concessão; documentos de patentes por tipo; documentos de patentes por status legal; principais requerentes; principais proprietários; principais classificações pelos códigos CPC (Cooperative Patent Classification) ou IPC (International Patent Classification); documentos de patentes por jurisdição; principais inventores; principais agentes e advogados; patentes mais citadas;
- trabalhos citados pelas patentes e indexados pela Lens, incluindo: artigos de periódicos; artigos de conferências; capítulos de livro; livros; dissertações e teses; relatórios técnicos.

Neste trabalho, foi utilizado o pacote R Bibliometrix (Aria & Cuccurullo, 2017, 2021), que é um *software* de código aberto para automatizar as etapas de análise e visualização de dados quantitativos e qualitativos. Depois de converter e enviar dados bibliográficos em R, o Bibliometrix possibilita uma análise descritiva do tema pesquisado. O Quadro 2 apresenta as bases que o pacote R Bibliometrix permite a importação dos metadados e os respectivos formatos.

Quadro 2 - Bases de dados bibliográficas e formatos dos metadados

Base	Formato
Web of Science	plaintext (‘.txt’), endnote (‘.ciw’) ou bibtex (‘.bib’)
Scopus	bibtex (‘.bib’) ou csv (‘.csv’)
Dimensions	csv (‘.csv’) ou excel (‘.xlsx’)
The Lens	csv export file (‘.csv’)
PubMed/MedLine	pubmed export file (‘.txt’)
Cochrane Library	plaintext (‘.txt’)

Fonte: Aria e Cuccurullo (2020).

A Figura 2 apresenta o processo de coleta, análise e visualização dos dados. É importante ressaltar que o pacote R-Bibliometrix apoia as principais etapas do fluxo de trabalho recomendado para o mapeamento científico. O fluxo tem início na coleta de dados nas bases e nos formatos mencionados no Quadro 1. Em seguida, os dados passam pelas etapas de criação das matrizes de documento e atributo para viabilizar o mapeamento semântico.

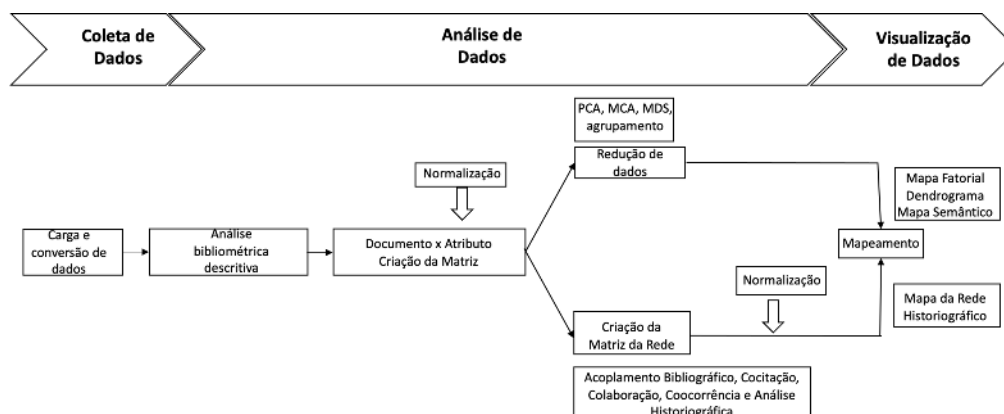


Figura 2 - Processo de coleta, análise e visualização de dados do pacote R-Bibliometrix.

Como mencionado anteriormente, o pacote R-Bibliometrix não permite a análise dos dados de patentes, porque os identificadores são diferentes. Assim, foi necessário fazer uma substituição dos identificadores de campos adequando ao formato dos artigos científicos, conforme mostrado no Quadro 3. É importante observar que alguns resumos podem estar no idioma de origem, havendo a necessidade de eliminação do registro ou a tradução para o idioma inglês. No identificador dos autores (*Authors*) é possível selecionar os inventores (*Inventors*), os proprietários (*Owners*) ou os requerentes (*Applicants*) das patentes.

A Figura 3 apresenta a página inicial da interface gráfica Biblioshiny, que implementa as funcionalidades do pacote R Bibliometrix. A interface possui as seguintes funcionalidades que podem ser executadas para a análise de patentes: importação de dados; visão geral, incluindo a produção científica anual e o gráfico com três campos; as fontes, com as principais jurisdições e a dinâmica dos registros por jurisdição; os autores (inventores,

requerentes e proprietários) incluindo os autores mais relevantes e a produção dos autores ao longo do tempo; os documentos incluindo as palavras mais frequentes, nuvem de palavras, dinâmica de palavras e tópicos de tendências; e a estrutura conceitual incluindo as redes de coocorrência, mapas temáticos, evoluções temáticas e análises fatoriais. Todas as funcionalidades que analisam palavras, extraem os dados dos códigos IPC, CPC e dos títulos e dos resumos.

Quadro 3 - Substituição dos identificadores de Patentes para Artigos.

Identificadores das Patentes	Identificadores dos Artigos
Lens ID	Lens ID
Title	Title
Jurisdiction	Source Country
Publication Date	Date Published
Publication Year	Publication Year
Application Number	ISSNs
Abstract	Abstract
Owners, Inventors ou Applicants	Author/s
URL	External URL
Document Type	Publication Type
Cited by Patents	Citing Patents
Extended Family Size	Citing Works Count
CPC Classifications/ IPC Classifications	Keywords

Detalhando um pouco mais a exploração da estrutura conceitual, que se baseia na coocorrência de palavras. Para as opções com as palavras-chave dos autores e palavras-chave de indexação, são geradas as redes de coocorrência de códigos CPC ou IPC, respectivamente, conforme estabelecido no Quadro 3. Deve-se optar por apenas uma das opções CPC ou IPC, porque o Bibliometrix processa apenas a rede de coocorrência dos códigos incluídos no atributo *Keywords*. Para a opção de títulos e resumos, é possível extrair os monogramas, os bigramas ou os trigramas. Um recurso interessante, que foi incluído na última versão do Bibliometrix, é a edição de texto com a possibilidade de incluir dois arquivos: um contendo uma lista de termos a serem removidos da análise e o outro com uma lista de sinônimos que serão fundidos em um único termo.

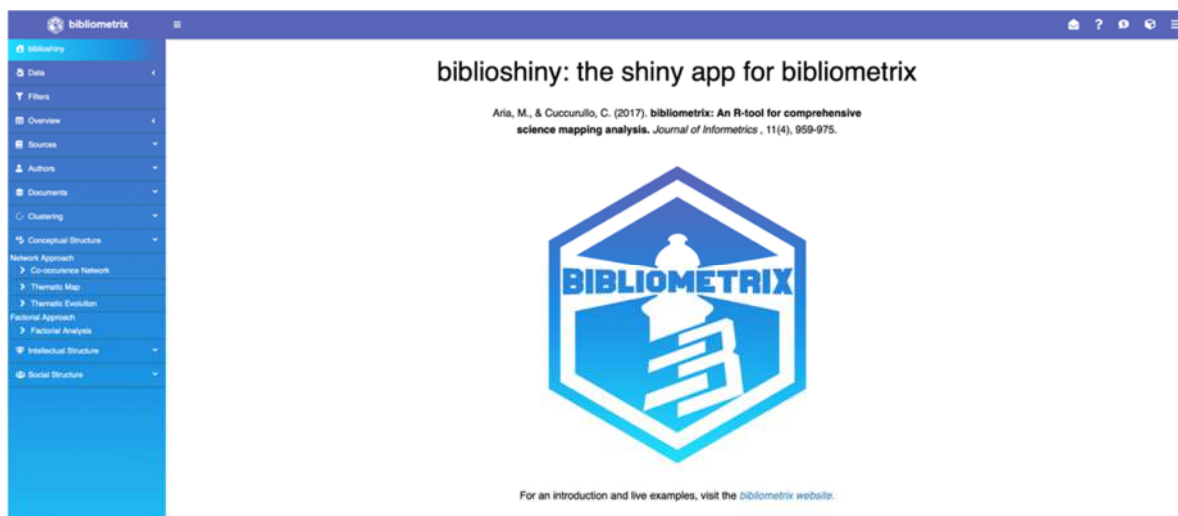


Figura 3 – Interface gráfica Biblioshiny que implementa as funcionalidades do pacote R Bibliometrix.

4. Aplicação da abordagem metodológica para a análise de patentes

Para exemplificar a abordagem apresentada na Figura 1, foi definido um tema de pesquisa atual – *virtual worlds*. O desenvolvimento de mundos virtuais, tanto em termos de características técnicas como em relação a uma gama cada vez maior de experiências relatadas pelos usuários, resultou em uma compreensão fragmentada na literatura do que é e do que não é um mundo virtual. Para complicar ainda mais este problema, a literatura de científica tem uma variedade de palavras-chave que são usadas para rotular a tecnologia: mundo virtual (*virtual world*); ambiente virtual (*virtual environment*); ambiente virtual multiusuário (*multiuser virtual environment*); jogo online (RPG) massivo-multiplayer (*massive multiplayer online game*(RPG)); mundo virtual imersivo (*immersive virtual world*); mundo imersivo (*immersive virtual world*); ambiente online imersivo (*immersive virtual environment*); ambiente de aprendizagem virtual 3D (*3D virtual learning environment*); mundos virtuais abertos (*open virtual worlds*); mundos simulados (*simulated worlds*); mundo virtual sério (*serious virtual world*); mundo virtual social (*social virtual world*); mundo virtual sintético (*synthetic virtual world*); ambiente de aprendizagem virtual (*virtual learning environment*), etc (Givan, 2018).

As características dos mundos virtuais influenciam a experiência dos usuários dentro desses espaços. Embora os parâmetros variem além de um ambiente simulado, as seguintes características gerais são necessárias para alcançar as experiências dos usuários e atender às exigências da estrutura: avatar, múltiplos usuários simultâneos, ferramentas de comunicação, persistência e representação do espaço. Portanto, há que se considerar dois aspectos da experiência do usuário que são componentes necessários de um mundo virtual: uma sensação de presença e de modelagem do mundo. Embora estas sejam experiências amplas, não relacionadas a nenhum contexto específico, elas podem ser aplicadas diversos contextos virtuais e destacar as principais diferenças experienciais entre os mundos físicos e virtuais.

Tendo combinado as propriedades determinantes, é agora possível afirmar a definição para um mundo virtual (VW) como:

um ambiente simulado onde: muitos agentes podem interagir virtualmente uns com os outros, agir e reagir às coisas, aos fenômenos e ao ambiente; os agentes podem ser zero ou muitos humanos, cada um representado por muitas entidades chamadas um "eu virtual" (um avatar) ou muitos agentes de software; toda ação/reação/interação deve acontecer em um ambiente virtual compartilhado em tempo real, espaço-temporal e não-pagável; o ambiente pode consistir de muitos espaços de dados, mas a coleta de dados Os espaços devem constituir um espaço de dados compartilhados, um fragmento persistente (Nevelsteen, 2017).

Søraker (2011) ressaltou que o conceito de virtualidade possui características relevantes, tais como simulação de computador, interatividade, indexação, múltiplos usuários e visão em primeira pessoa, e sugeriu as seguintes definições:

- virtual x: interativo, simulado por computador x (ou, x tornado possível pela simulação interativa por computador);
- ambiente virtual: ambiente interativo, simulado por computador, indexado;
- mundo virtual: ambiente interativo, simulado por computador, indexado, multiusuário;
- realidade virtual: ambiente interativo, simulado por computador, experimentado a partir de uma visão em primeira pessoa.

A relação entre estes conceitos pode ser melhor ilustrada na Figura 4. O objetivo é realizar um estudo exploratório sobre o tema, verificando o que está patenteado e o que está sendo requerido. Para uma análise mais abrangente, foi realizada uma pesquisa na base Lens sobre as patentes submetidas e concedidas sobre *virtual worlds*. A pesquisa retornou 6446 documentos, sendo 1909 patentes concedidas, 73 modelos de utilidade e 4464 patentes submetidas.

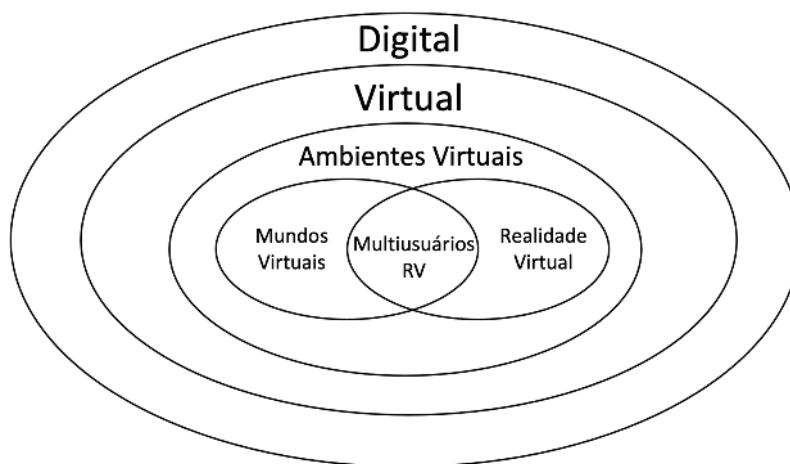


Figura 4 - Relação e interseções entre as diferentes categorias de virtualidade (Søraker, 2011).

A Figura 5 apresenta a evolução da produção de patentes no período de 1990 a 2022. O gráfico foi gerado pelo pacote R-Bibliometrix. É possível observar que o pico ocorreu em 2020 com 490 documentos publicados. A taxa de crescimento anual calculada pelo pacote R-Bibliometrix é de 18,96%.

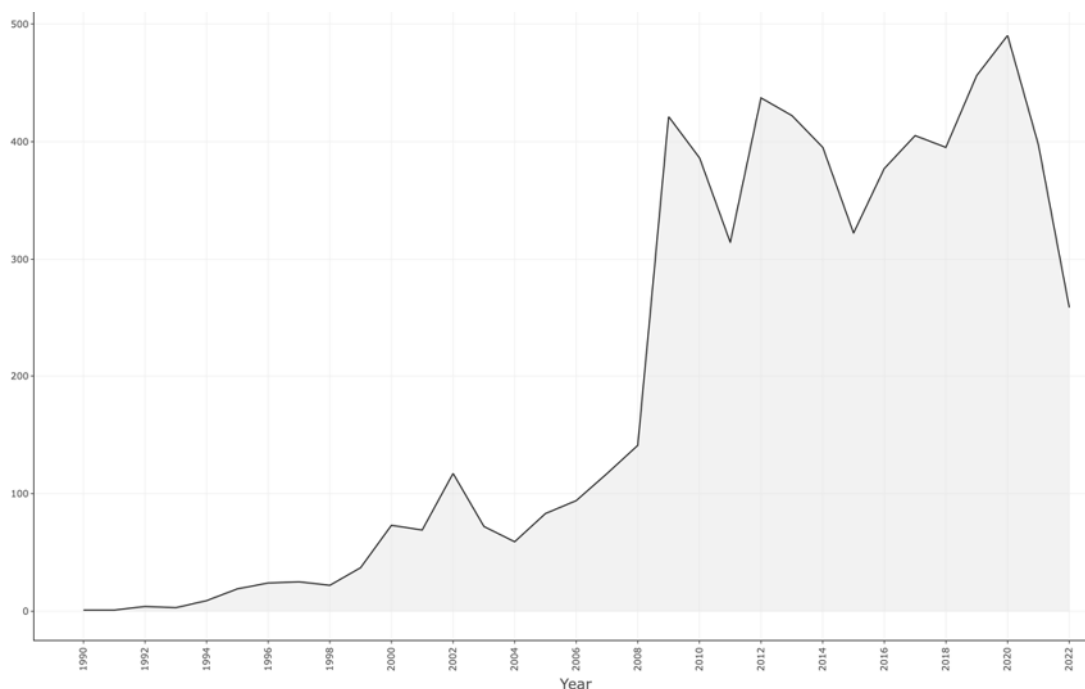


Figura 5 – Evolução anual da produção de patentes.

A base Lens permite extrair alguns indicadores de desempenho relativos à jurisdição, aos detentores das patentes e aos inventores. A Tabela 1 apresenta os principais escritórios de

propriedade intelectual onde foram registradas as patentes. Observa-se que o escritório dos Estados Unidos concentra o maior número de documentos registrados, seguido pela China e pelo World Intellectual Property Organization.

Tabela 1 – Número de documentos por país.

Jurisdiction	Número de Documentos
United States	3860
China	708
WO - WIPO	620
Japan	362
European Patents	331
Korea, Republic of	250
Canada	106
Australia	65
United Kingdom	60
Russia	16
Germany	14
Taiwan	11

Fonte: base Lens.

A Tabela 2 apresenta as principais companhias detentoras de patentes. A IBM possui a maior quantidade documentos, seguida pela Searete e Microsoft.

Tabela 2 – Principais detentores de patentes.

Owner Name Exact	Número de Documentos
IBM	622
Searete	192
Microsoft	278
Sony	183
Nintendo	164
Activision Publishing INC	113
Disney Enterprises INC	97
The Invention Science Fund I	94
Uthervers Digital INC	91
Kyndryl INC	74
Google	64
Samsung Electronics Co LTD	59

Fonte: base Lens.

A Tabela 3 apresenta os 20 principais inventores com os respectivos vínculos institucional. A empresa Searete possui o maior número de inventores, em um total de oito, seguido pela Samsung com quatro, IBM com três e Magic Leap e Ganz com dois cada.

Tabela 3 – Principais inventores e vínculo institucional

Inventor	Número de Documentos	Principal Vínculo Institucional
Royce A Levien	204	Searete
Mark A Malamud	193	Searete

Robert W Lord	182	Searete
Rick A Hamilton II	118	IBM
Howard Ganz	109	Ganz
Seung Ju Han	90	Samsung
Edward K Y Jung	90	Searete
Clifford A Pickover	90	IBM
Eric C Leuthardt	88	Searete
Jae Joon Han	87	Samsung
Lowell L Wood Jr	79	Searete
Samuel A Miller	73	Magic Leap
Bang Won Chul	71	Samsung
Kim Do Kyoon	70	Samsung
Rinaldo John D Jr	69	Searete
Abovitz Rony	64	Magic Leap
Cevat Yerli	63	TMRW Found IP & Holding Sarl
Karl Joseph Borst	55	Ganz
Brian M O'connell	54	IBM
John D Rinaldo Jr	54	Searete

Fonte: base Lens.

A Figura 6 apresenta a produção dos inventores ao longo do período de 2006 a 2022. O gráfico foi gerado pelo pacote R-Bibliometrix.

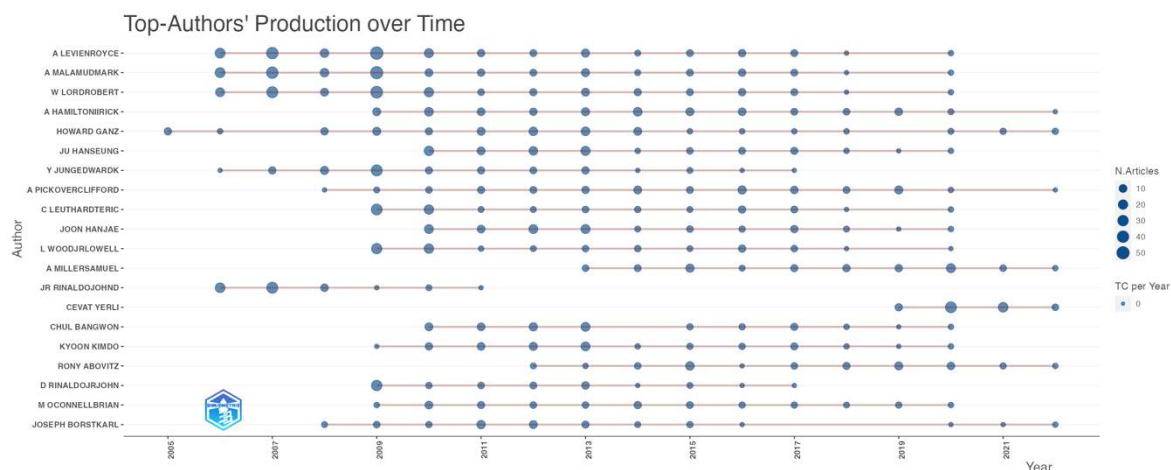


Figura 6 – Produção dos inventores no período de 2006 a 2022.

Para analisar a estrutura tecnológica, o pacote R-Bibliometrix oferece vários recursos para extração de palavras das palavras-chave, dos títulos e dos resumos. A Figura 7 apresenta as palavras mais frequentes extraídas dos bigramas dos resumos. A palavra *virtual world*, que deu origem à pesquisa, destaca-se com 9949 ocorrências, seguida de *real world* com 2159, *virtual object* com 1304, *virtual reality* com 947 e *virtual environment* com 818.

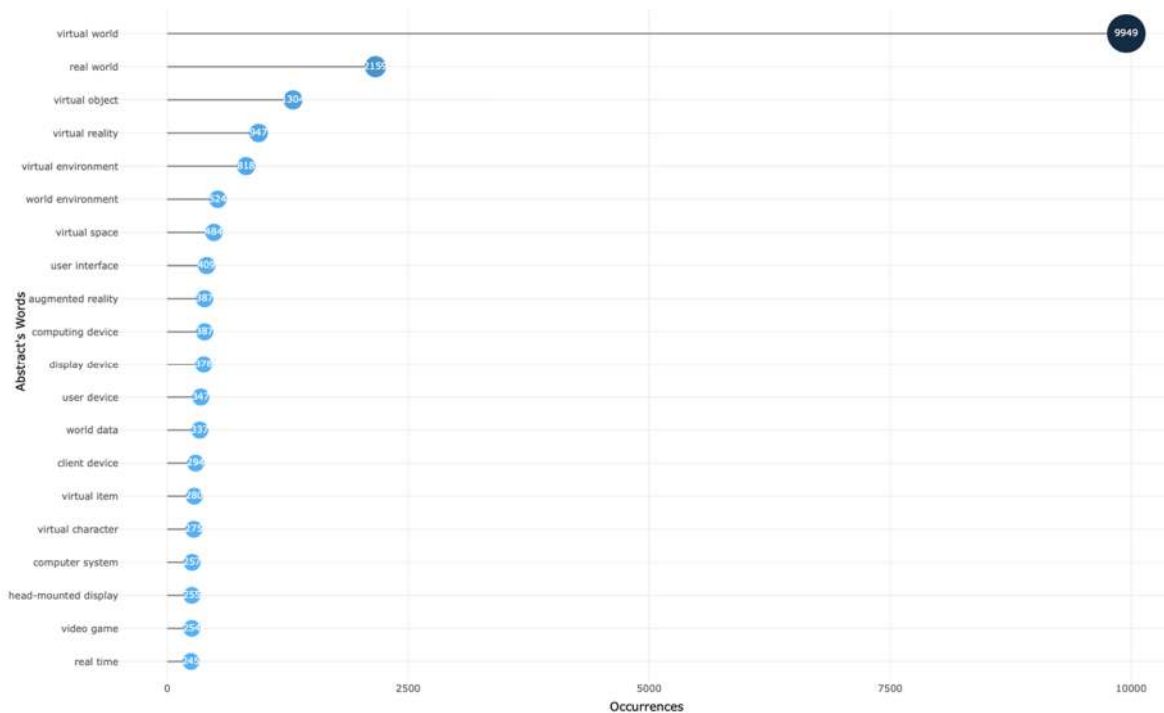


Figura 7 - Palavras mais frequentes extraídas dos bigramas dos resumos.

A pesquisa identificou 3972 códigos CPC – subgrupo, 619 - grupo, 166 - subclasses e 58 classes. A Tabela 4 apresenta os códigos de classes CPC mais frequentes, com as ocorrências e as respectivas subclasses. Os dados foram extraídos da função Most Frequent, aplicando, respectivamente, a lista de sinônimos que reduz os códigos de subgrupo para classe e para subclasse. A classe G06 – computação, cálculo e contagem – apresenta a maior quantidade de ocorrências com 30583, seguida pela classe A63 – esportes, jogos e diversões. Este resultado é consistente, considerando o conceito central dos mundos virtuais.

Tabela 4 – Códigos de classes mais frequentes.

Classe CPC	Ocorrências	Subclasses CPC
g06 – computação; cálculo; contagem	30583	g06f (13079), g06t (8114), g06q (6971), g06n (1181), g06v (856), g06k (380) e g06e (2)
a63 – esportes, jogos e diversões	29800	a63f (28097), a63b (1133), a63h (412), a63g (122), a63c (24), a63j (6), a63k (4), a63d (2)
h04 – técnica de comunicação elétrica	11256	h04l (5181), h04n (4198), h04m (743), h04w (559), h04s (330), h04r (137), h04h (40), h04q (35), h04b (21), h04j (12)
g02 - ótica	4199	g02b (4037), g02f (160), g02c (2)
a61 - ciência médica ou veterinária; higiene	3719	a61b (2625), a61m (812), a61h (198), a61n (54), a61f (12), a61k (8), a61c (7), a61g (2), a61p (1)
g09 - educação; criptografia; exposição; publicidade; lacres	1702	g09g (1002), g09b (683), g10h (78), g09c (9), g09f (8)
g07 - dispositivos de verificação	982	g07f (950), g07c (32)
g16 - tecnologia da informação e comunicação (TIC) especialmente adaptada para campos de aplicação específicos	720	g16h (625), g16z (53), g16b (32), g16y (10)

g01 - medição; testes	547	g01c (238), g01s (185), g01b (29), g01n (26), g01d (18), g01m (18), g01w (13), g01p (9), g01g (7), g01j (4)
g05 - métodos, sistemas e aparelhos para o controle, em geral	532	g05b (342), g05d (184), g05g (6)
b60 - veículos em geral	494	b60r (222), b60w(208), b60k (48), b60q (4), b60s (4), b60t (4), b60y (3), b60l (1)

Fonte: elaborada pelos autores.

Para a identificação dos tópicos de tendências, foi utilizada a função Trend Topics do menu Documents e a lista de sinônimos para reduzir os códigos CPC – subgrupo para subclasse. Os parâmetros foram ajustados para mínimo de 1 frequência do código e 20 códigos por ano. A Tabela 5 apresenta as tendências tecnológicas para os códigos CPC de subclasses. Os códigos C03C e G02C são as tecnologias emergentes. O código C03C possui duas patentes submetidas pela Microsoft Technology Licensing LLC, com registros no USPTO (Sakari et al, 2022a) e WIPO (Sakari et al, 2022b), referente a revestimentos dielétricos para melhorar a uniformidade da exibição e a segurança da luz em um combinador óptico. O código G02C possui duas patentes concedidas pelo USPTO (Navee, 2021, 2022), referente a sistemas e métodos para operação assistida por computador.

Tabela 5 – Tendências tecnológicas

Subclasse	Freq	Ano_q1	Ano_med	Ano_q3
c03c	12	2022	2022	2022
g02c	2	2021	2022	2022
h02j	14	2017	2022	2022
g16y	10	2021	2021	2022
f16m	2	2021	2021	2021
b23k	1	2021	2021	2021
g03f	1	2021	2021	2021
b60s	4	2020	2021	2022
b60t	4	2020	2021	2022
h02k	15	2020	2021	2021
a61c	7	2020	2021	2021
y02a	6	2020	2021	2021
h04s	330	2019	2021	2021

Fonte: elaborada pelos autores.

As redes de palavras permitem analisar a estrutura conceitual, evidenciando as ligações entre conceitos através de coocorrências de termos. O pacote R Bibliometrix permite obter a rede de coocorrência de palavras-chave e sua relação com o mapa temático. As palavras-chave representam conceitos, cuja densidade e centralidade podem ser usadas na categorização e no mapeamento conceitual em um diagrama bidimensional. As comunidades detectadas podem ser representadas por duas medidas (Callon, Courtial, & Laville, 1991), ou seja, o grau de relevância (centralidade de Callon) e o grau de desenvolvimento (densidade da Callon). A centralidade de Callon mede a intensidade dos vínculos entre uma determinada comunidade e as demais. O valor pode ser representado como uma medida da importância de um tema em todo o corpus. A densidade da Callon mede a força interna da comunidade. Este valor pode ser representado como uma medida do desenvolvimento do tema. Com base nessas duas medidas, os temas de pesquisa podem ser mapeados em um diagrama estratégico

bidimensional com quatro quadrantes: (1) quadrante superior direito: temas motores; (2) quadrante inferior direito: temas básicos; (3) quadrante inferior esquerdo: temas emergentes ou em desaparecimento; (4) quadrante superior esquerdo: temas de nicho ou muito especializados.

Geralmente, as palavras-chave no quadrante superior direito, conhecidas como temas motores, são bem desenvolvidos e importantes para estruturar um campo de pesquisa. Os temas no quadrante superior esquerdo são de importância apenas marginal para o campo com laços internos bem desenvolvidos, mas com laços externos sem importância. Os temas no quadrante inferior esquerdo são pouco desenvolvidos e marginais. Eles representam principalmente temas emergentes ou em desaparecimento. Temas transversais e básicos estão contidos no quadrante inferior direito, e são importantes, mas não estão desenvolvidos.

Neste trabalho, foram gerados dois mapas temáticos: códigos CPC e bigramas dos resumos. A Figura 8 apresenta o mapa temático das subclasses do código CPC. Foram obtidas 5 comunidades tecnológicas:

- a63f - Saúde, salvamento e recreação - jogos de cartas, mesa ou roleta; jogos em recintos fechados usando pequenas peças móveis para jogo; videogames; jogos não incluídos em outro local;
- g06f - Computação, cálculo e contagem - processamento elétrico de dados digitais;
- g09b - Educação; criptografia; apresentação visual; anúncios; logotipos - aparelhos educativos ou de demonstração; aparelhos para ensino ou comunicação com os cegos, surdos ou mudos; modelos; planetários; globos; mapas; diagramas;
- g06k - Computação, cálculo e contagem - leitura de dados gráficos; apresentação de dados; suporte de dados; manipulação de transportes de dados;
- g02f – Óptica - dispositivos ou disposições ópticos para controle de luz pela modificações de propriedades ópticas do meio dos elementos envolvidos nos mesmos; óptica não linear; mudança de frequência da luz; elementos ópticos lógicos; conversores ópticos analógicos/digitais.

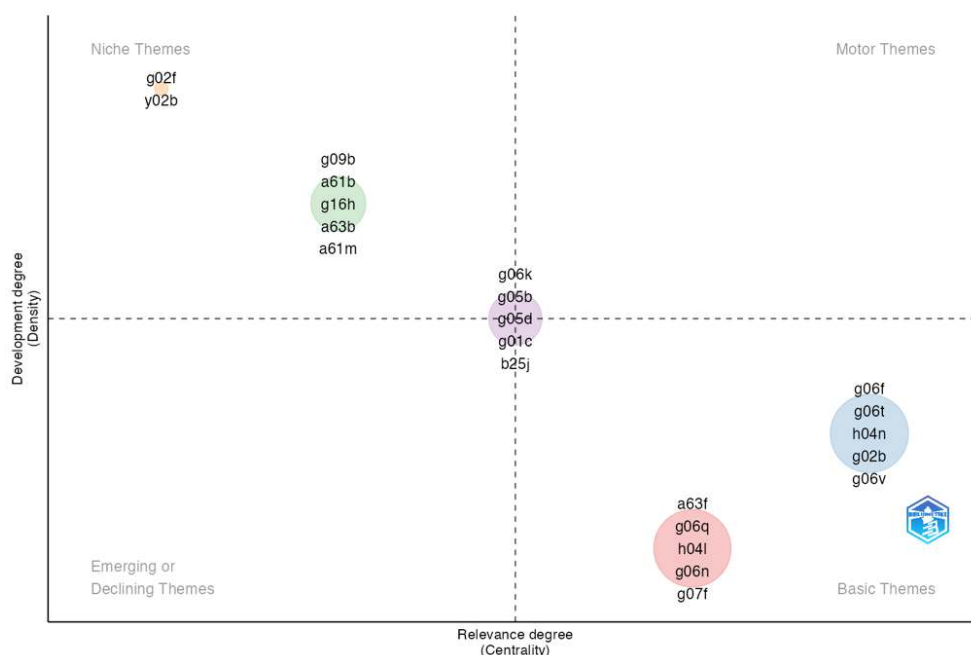


Figura 8 – Mapa temático das subclasses do código CPC.

A Tabela 6 apresenta o detalhamento das comunidades acima e o posicionamento no mapa temático. É possível selecionar documentos que contribuíram com cada comunidade. O pacote R-Bibliometrix gerou uma tabela com os 6446 documentos com a identificação da comunidade e a sua influência utilizando a métrica do PageRank (Brin & Page, 1998).

Tabela 6 – Comunidades do mapa temático das subclasses dos códigos CPC.

Comunidade	Tema	Frequência	Códigos
a63f	Básico	6348	a63f (2572), g06q (1559), h04l (1118), g06n (373), g07f (265), h04m (116), h04w (209), a63h (124), h04h (12)
g06f	Básico	6917	g06f (2867), g06t (2042), h04n (682), g02b (489), g09g (213), g06v (258), h04s (47), g10l (95), y10s (64), h04r (33), a63g (42), g10h (13), g03b (29), g11b (24), y02d (19)
g09b	Nicho	929	a61b (216), a63b (113), a61m (38), g09b (295), g16h (184), a61h (22), f41g (13), h04q (12), g01b (20), h04a61b (16), a63b (113), a61m (38), g09b (295), g16h (184), a61h (22), f41g (13), h04q (12), g01b (20), h04b (16)
g06k	Central	799	g06k (169), g05b (86), g01c (74), g08b (27), b60r (41), b60w (39), g01s (53), g05d (80), b25j (60), g08g (38), b64c (23), y02p (31), g16z (22), b64d (17), g07c (23), h05b (16)
g02f	Nicho	26	g02f (13), y02b (13)

Fonte: elaborada pelos autores.

A Figura 9 apresenta o mapa temático dos bigramas do resumo. As palavras extraídas pelo pacote R-Bibliometrix foram aglutinadas nas seguintes comunidades: *virtual world*; *virtual environment*; *world environment*; *user interface*; *computing device*; *world system*; *computer system*; *real time*; *computer program*; *processing unit*; *reality environment*; *user avatar*; *gaming system*; *computer implemented*; *location information*; *data representing*; *reality game*; *electronic game*.

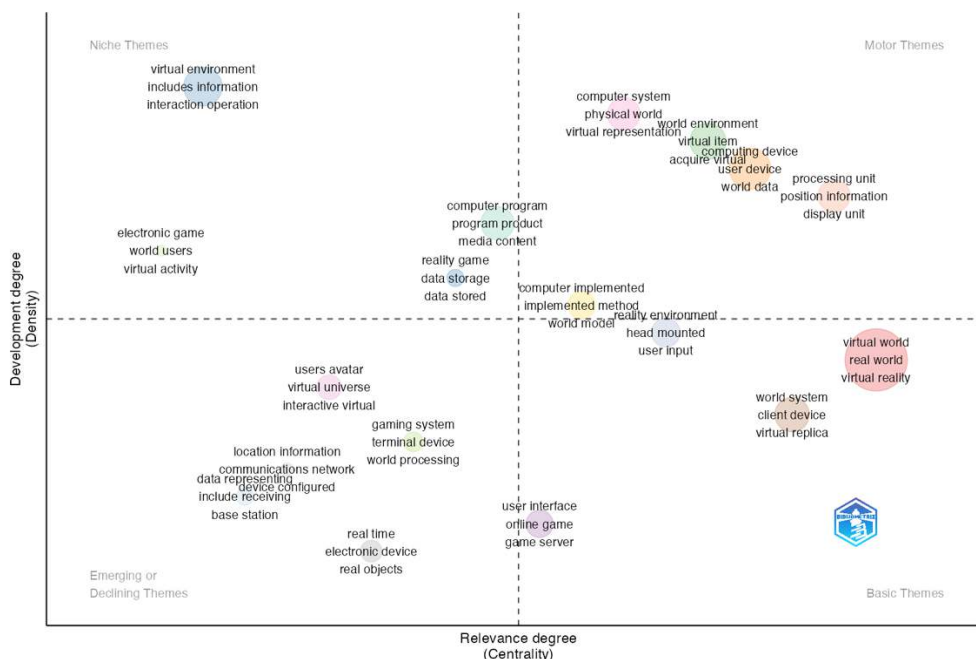


Figura 9 – Mapa temático dos bigramas dos resumos das patentes.

A Tabela 7 apresenta os resultados do mapa temático obtido pela extração dos bigramas dos resumos das patentes. Como foram obtidas 268 palavras-chave, foram incluídas nesta tabela apenas as que tiveram maiores frequências. Um resultado interessante é que as comunidades relacionadas à computação são temas motores, que são bem desenvolvidos e importantes para estruturar um campo de pesquisa. Isso é coerente com a definição de Callon, Courtial e Laville (1991), com os graus de relevância e de desenvolvimento mais altos.

Tabela 7 – Comunidades do mapa temático dos bigramas extraídos dos resumos das patentes.

Comunidade	Tema	Freq.	Principais Palavras-chave
virtual world	Básico	11083	virtual world (4141), real world (1080), virtual object (458), virtual reality (478), virtual space (177), augmented reality (245), display device (213), video game (138)
virtual environment	Nicho	1318	virtual environment (431), includes information (46), interaction operation (39), operating virtual (39), world interaction (35), information systems (36)
computing device	Motor	1631	computing device (216), user device (191), world data (135), computer network (108), processing circuitry (95), social network (76)
world environment	Motor	946	world environment (276), virtual item (119), virtual products (100), world activities (48), acquire virtual (74)
world system	Básico	938	client device (114), virtual replica (82), world system (116), persistent virtual (75), real object (60), physical location (57), computer-implemented method (56)
computer system	Motor	764	computer system (142), physical world (110), virtual representation (93), entertainment system (50), registration code (67), user access (50)
processing unit	Motor	736	position information (81), processing unit (96), map information (37), display data (39), virtual display (47), image processing (50)
computer program	Nicho	704	computer program (212), program product (200), media content (65), sensory experience (65), artificial sensory (64), bioactive agent (47)
reality environment	Básico	532	reality environment (88), head mounted (69), mounted display (61), physical environment (43)
user interface	Básico	465	user interface (244), online game (81), game server (65), game character (36), data representative (39)
computer implemented	Motor	440	world model (44), three-dimensional space (43), computer implemented (53), implemented method (50)
user avatar	Emergente ou declínio	366	virtual universe (60), user avatar (81), interactive virtual (54), object based (47), user viewing (43)
real time	Emergente ou declínio	276	real time (191), electronic device (47), real objects (38)
gaming system	Emergente ou declínio	232	portable device (35), terminal device (43), gaming system (48), world processing (39)
reality game	Nicho	183	reality game (41), parallel reality (34), data storage (38), game data (32), data stored (38)
data representing	Emergente ou declínio	137	data representing (72), base station (33), interactive media (32)
location information	Emergente ou declínio	120	location information (52), communications network (34), device configured (34)
electronic game	Nicho	103	electronic game (36), virtual activity (32), world users (35)

Fonte: elaborada pelos autores.

Outra funcionalidade importante do pacote R-Bibliometrix é a evolução temática, que auxilia na interpretação da estrutura conceitual. Esta estrutura é frequentemente usada para compreender a evolução das tecnologias, utilizando os códigos CPC, ou os tópicos

abordados pelos inventores, pela extração dos monogramas, bigramas ou trigramas dos títulos ou dos resumos, e identificar quais são os conceitos ou as tecnologias mais importantes e mais recentes. Dividir o tempo em diferentes períodos e comparar as estruturas conceituais e tecnológicas é útil para analisar a evolução dos tópicos ao longo do tempo. A Figura 10 apresenta a evolução da estrutura tecnológica das subclasses dos códigos CPC divididos em cinco períodos: 1990-2009; 2010-2013; 2014-2016; 2017-2019; e 2020-2022. Os recortes foram definidos a partir dos picos identificados na produção científica anual, apresentada anteriormente na Figura 5.

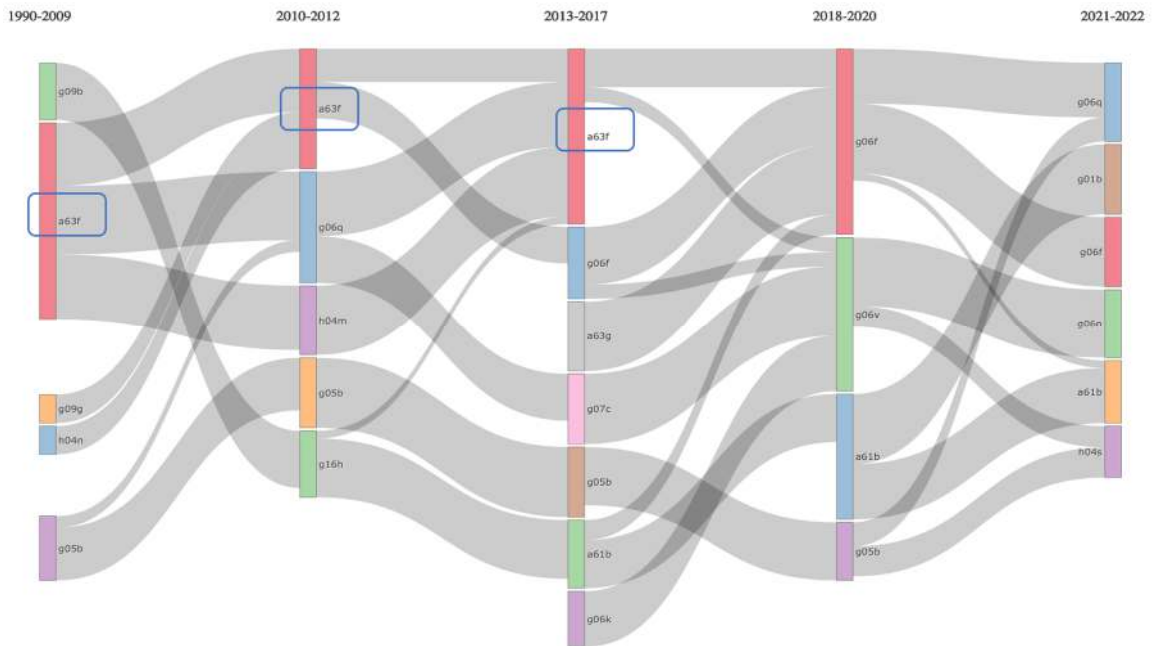
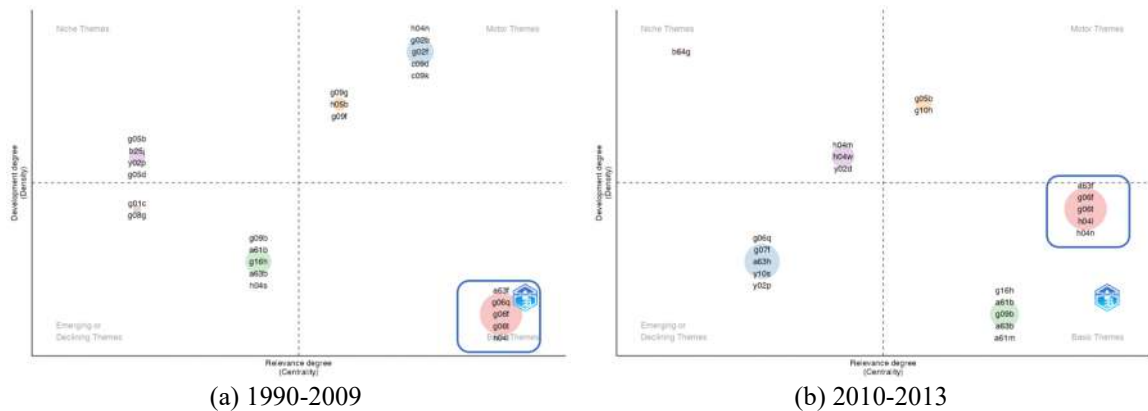


Figura 10 –Evolução temática da estrutura tecnológica a partir das subclasses dos códigos CPC das patentes.

A subclasse A63F, que contempla diversos tipos de jogos é a que possui maiores ocorrências nos três primeiros períodos. Apesar de não aparecer na Figura 10 nos últimos dois períodos, a subclasse continua sendo utilizada nas classificações das patentes. A Figura 11 apresenta os mapas temáticos de cada período. Observa-se que a subclasse A63F é um tema básico constante em todos os períodos.



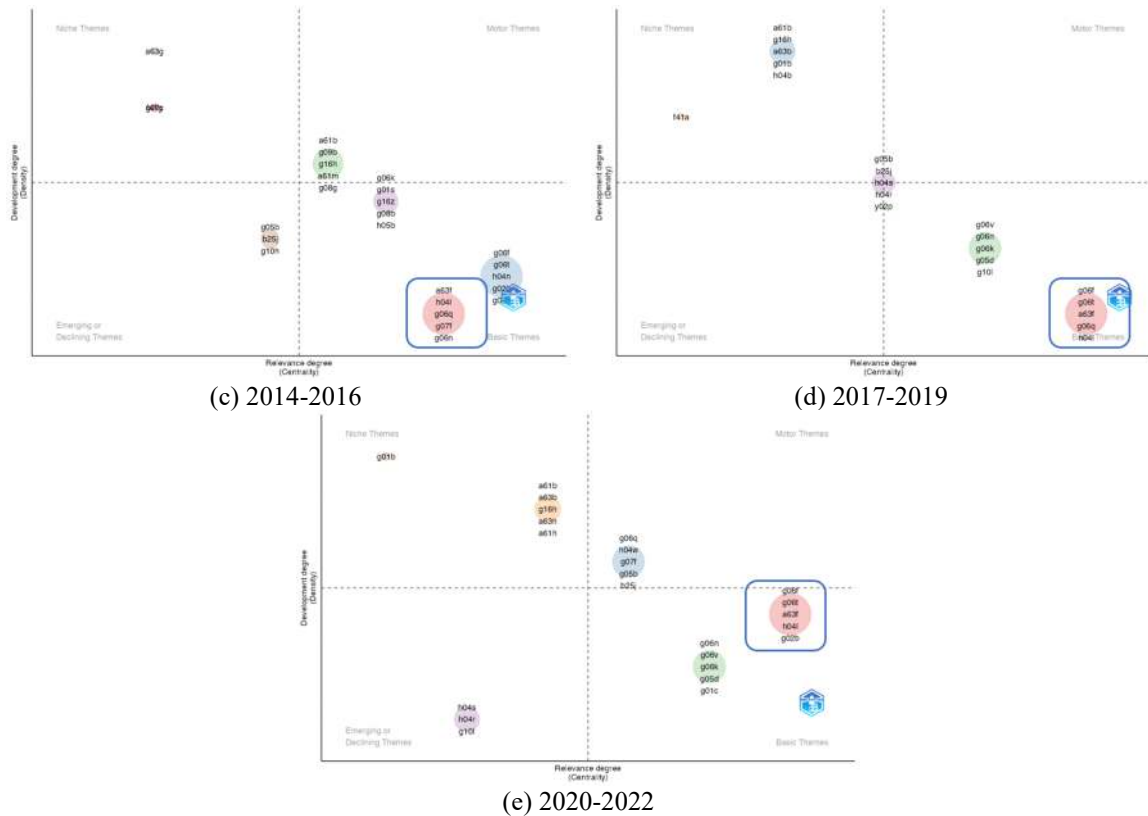


Figura 11 – Mapas temáticos da evolução da estrutura tecnológica por período.

A Figura 12 apresenta a evolução temática da estrutura tecnológica a partir os bigramas extraídos dos resumos. O tema da pesquisa se destaca em todos os períodos. Na análise do fluxo entre os períodos é possível verificar as rotas conceituais da tecnologia em cada período. Por exemplo, no período de 1990 a 2009, o conceito tecnológico de mundos virtuais, como um tema básico, contribui com quase todos do período seguinte.

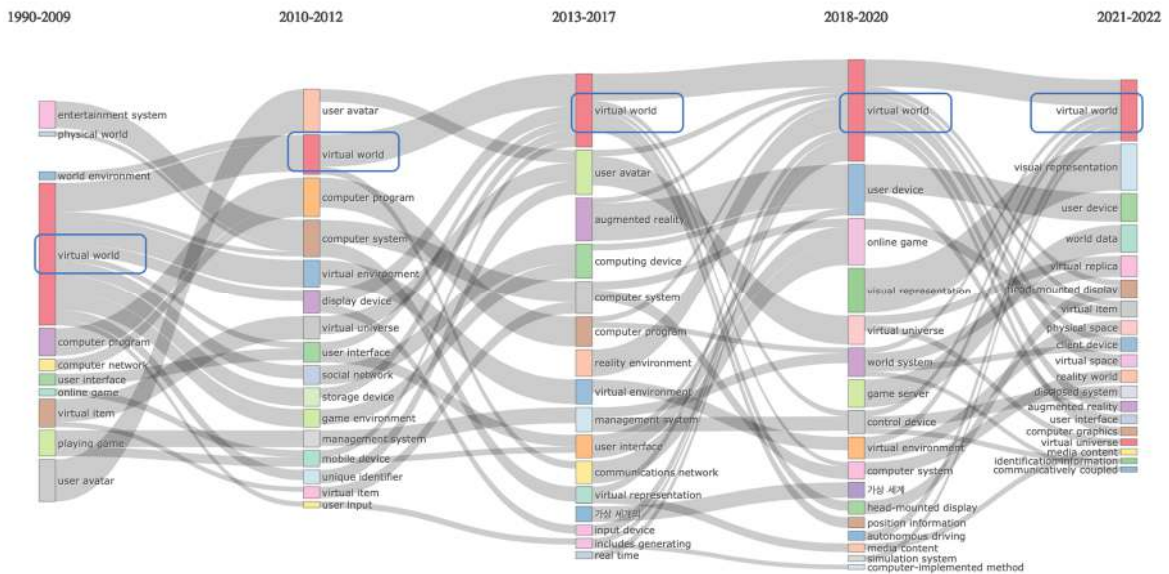


Figura 11 –Evolução temática da estrutura tecnológica a partir dos bigramas extraídos dos resumos das patentes.

Conclusões

O objetivo deste artigo foi a apresentação de uma abordagem de análise bibliométrica para interpretar documentos recuperados em pesquisas realizadas em bases de patentes. A abordagem proposta parte do desenho da pesquisa, com a definição dos objetivos e de sua abrangência, que é essencial para as escolhas que serão necessárias nas próximas etapas.

Apesar do pacote R-Bibliometrix não permitir a análise dos registros recuperados na pesquisa, foi possível fazer uma conversão dos identificadores dos metadados das patentes para o formato adequado para a importação seguindo o padrão dos identificadores dos artigos. Há uma limitação, porque nem todos os dados são exportados pela base, como é o caso dos documentos citados pelas patentes. Assim, não foi possível explorar a estrutura intelectual das informações registradas nas patentes, devido às limitações referentes às referências citadas pelas patentes.

A base Lens permite identificar uma família de patentes, que é um grupo de invenções que estão relacionadas a um documento específico. Nesse sentido, é possível aprofundar a interpretação de uma tecnologia ou inovação através da análise qualitativa de todos os documentos de patentes de uma família.

Para exemplificar a abordagem foi realizada na base Lens com a expressão de busca – virtual world*. Após a aplicação de filtros para delimitar o resultado ao escopo da pesquisa, foram recuperados 6446 documentos, sendo 1909 patentes concedidas, 73 modelos de utilidade e 4464 patentes submetidas. A aplicação da abordagem permitiu identificar alguns indicadores de desempenho e traçar um panorama tecnológico do tema, incluindo a evolução anual da produção de patentes, os quantitativos de documentos por jurisdição, os principais detentores de patentes, os principais inventores com os respectivos vínculos institucionais, a produção dos inventores no período de 2006 a 2022, as palavras-chave mais frequentes extraídas dos resumos, os códigos das classes CPC mais frequentes, as tendências tecnológicas a partir das subclasses CPC, os mapas temáticos e as evoluções da estrutura tecnológica das subclasses CPC e dos bigramas extraídos dos resumos.

Como perspectiva futura, sugere-se o aprofundamento do entendimento das funcionalidades do pacote R-Bibliometrix, particularmente quanto à interpretação dos identificadores que são exportados pela base Lens para uma patente e o seu correspondente na exportação de artigos. Isso possibilitará novas análises dos dados qualitativos registrados nas patentes. Outra possibilidade é explorar a rede coocorrência de códigos CPC para realizar o refinamento da pesquisa.

Agradecimentos

Agradecemos à Universidade Católica de Brasília e à Universidade de Aveiro pela parceria na elaboração deste artigo. Este projeto foi apoiado pelo programa PPI Softex, Convênio nº 0200-11/2021/SOFTEX/UCB/ResidênciaTIC4, financiado pelo Ministério da Ciência, Tecnologia e Inovações com recursos da Lei nº 8.248, de 23 de outubro de 1991.

Referências Bibliográficas

- Abraham, B. & Morita, S. (2001). Innovation assessment through patent analysis. *Technovation*, 21 (4), 245-252.
- Almeida, F. L. C., Castro, M. P. J., Travália, B. M. & Forte, M. B. S. (2021). Trends in lipase immobilization: Bibliometric review and patent analysis. *Process Biochemistry*, 110, 37–51.
- Aria, M., & Cuccurullo, C. (2017). Bibliometrix: An R-tool for comprehensive science mapping analysis. *Journal of Informetrics*, 11 (4), 959-975.
- Aria, M., & Cuccurullo, C. (2020). *A brief introduction to bibliometrix*. Disponível em: https://www.bibliometrix.org/vignettes/Introduction_to_bibliometrix.html. Acesso: 10 jul 2020.
- Aria, M., & Cuccurullo, C. (2021). *Package 'bibliometrix'*. Disponível em: <https://cran.r-project.org/web/packages/bibliometrix/bibliometrix.pdf>. Acesso: 10 jul 2021.
- Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An Open Source Software for Exploring and Manipulating Networks. In: *Proceedings of the Third International ICWSM Conference*, p. 361-362.
- Bengisu, M. & Nekhili, R. (2006). Forecasting emerging technologies with the aid of science and technology databases. *Technological Forecasting & Social Change*, 73, 835-844.
- Brin, S. & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*. 30 (1-7), 107–117.
- Chen, Y. H., Chen, C. Y., & Lee, S. C. (2010). Technology forecasting of new clean energy: The example of hydrogen energy and fuel cell. *African Journal of Business Management*, 4 (7), 1372–1380.
- Clarivate. *Derwent Innovations Index*, 2021. Disponível em: <https://webofscience.help.clarivate.com/en-us/Content/derwent/derwent.htm>. Acesso em 03 out. 2022.
- CORE. The world's largest aggregator of open access research papers. 2022. Disponível em: <https://core.ac.uk>. Acesso em: 15 mar. 2022.
- Daim, T. U., Rueda, G., Martin, H. & Gerdtsri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73(8), 981–1012.
- Daim, T. U., Van Blommestein, K., Islam, N., Ozcan, S., Hillegas, J. & Estep, J. (2014). Integrating data mining into technology roadmapping. In: *2014 Portland International Center for Management of Engineering and Technology, PICMET 2014*, 2972–2985.
- European Patent Office. (2022). *Cooperative Patent Classification (CPC)*, 2022. Disponível em: <https://www.cooperativepatentclassification.org/home>. Acesso em: 15 mar. 2022.

Girvan, C. (2018). What is a virtual world? Definition and classification. *Education Technology Research and Development*, 66, 1087–1100.

Kauffman, J., Kittas, A., Bennett, L., & Tsoka, S. (2014). DyCoNet: A Gephi Plugin for Community Detection in Dynamic Complex Networks. *PLoS ONE*, 9(7), Art. e101357.

Klavans, R., & Boyack, K. W. (2006). Identifying a better measure of relatedness for mapping science. *Journal of the American Society of Information Science*, 57(2), 251–263.

Lai, K.-K., Lin, M.-L., Chang, S.-B. & Hsu, C.-F. (2007). The study of taxonomy and evolutionary trends of relevant literatures on patent analysis. In: PICMET '07 - Portland International Center for Management of Engineering and Technology - Management of Converging Technologies, 22–30.

Lens. (2022). *Search, Analyze and Manage Patent and Scholarly Data*, 2022. Disponível em: <https://www.lens.org>. Acesso em: 10 jan. 2022.

Liu, X. & Yu, X. (2017). Patent analysis for guiding technology transfer from EU/EEA to China: The case of CO2 compressor in CCUS cooperation. In: 2016 Portland International Conference on Management of Engineering and Technology, PICMET 2016, p. 1659–1671.

Morris, S., DeYong, C., Wu, Z., Salman, S., Yemenu, D. (2002). DIVA: a visualization system for exploring document databases for technology forecasting. *Computers and Industrial Engineering*, v. 43, n. 4, p. 841–862.

Naveen, K. (2021). *Systems and methods for computer assisted operation* (U. S. Patent No. 11,173,391 B1). Washington, DC: U.S. Patent and Trademark Office. <https://lens.org/069-143-860-288-246>

Naveen, K. (2021). *Systems and methods for computer assisted operation* (U. S. Patent No. 11,173,391 B1). Washington, DC: U.S. Patent and Trademark Office. <https://lens.org/033-195-717-414-826>

Nevelsteen, K. J. L. (2017). Virtual world, defined from a technological perspective and applied to video games, mixed reality, and the Metaverse. *Computer Animation and Virtual Worlds*, 29 (1), e1752.

Norton, M. J. (2010). *Introductory Concepts in Information Science*. 2nd Ed. New Jersey: Information Today Inc.

ORCID. *Connecting research and researchers*. 2022. Disponível em: <https://orcid.org>. Acesso em 10jan2022.

Piwowar, H., Priem, J., Larivière, V., Alperin, J. P., Matthias, L., Norlander, B., Farley, A., West, J., Haustein, S. (2018). The state of OA: a large-scale analysis of the prevalence and impact of Open Access articles. *PeerJ*, 6, e4375. <https://doi.org/10.7717/peerj.4375>

Sakari, V. T. H, Pietari, T., & Tapio, T. J. K. (2022a). Dichroic coatings to improve display uniformity and light security in an optical combiner (U. S. Patent No. US 2022/0146827 A1). Washington, DC: U.S. Patent and Trademark Office.

Sakari, V. T. H, Pietari, T., & Tapio, T. J. K. (2022b). Dichroic coatings to improve display uniformity and light security in an optical combiner (WIPO Patent No. 2022/098447 A1). Geneva, Switzerland: World Intellectual Property Organization.

Sheikh, N. J. & Sheikh, O. (2017). Forecasting of biosensor technologies for emerging point of care and medical IoT applications using bibliometrics and patent analysis. In: 2016 Portland International Conference on Management of Engineering and Technology, PICMET 2016, 3082–3093.

Søraker, J. H. (2011). Virtual entities, environments, worlds and reality: suggested definitions and taxonomy. In: Charles Ess & May Thorseth (eds.), *Trust and Virtual Worlds*. New York: Peter Lang.

van Eck, N. J., & Waltman, L. (2014). Visualizing bibliometric networks. In: Ding, Y., Rousseau, R., Wolfram, D. (Eds.). *Measuring scholarly impact: methods and practice*. New York: Springer.

van Eck, N. J., & Waltman, L. (2021). *VOSviewer manual*. Leiden: Universiteit Leiden.

Waltman, L., & van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology*, v. 63, n. 12, 2378–2392.

Wantanabe, C., Tsuji, Y. & Brown, C. (2001). *Patent statistics: deciphering a 'real' versus a 'pseudo' proxy of innovation*. *Technovation*, 21(12), 783-790.

World Intellectual Property Organization. IPC Publication. 2022. Disponível em: <https://ipcpub.wipo.int/?notion=scheme&version=20220101&symbol=none&menulang=en&lang=en&viewmode=f&fipcpc=no&showdeleted=yes&indexes=no&headings=yes¬es=yes&direction=o2n&initial=A&cwid=none&tree=no&searchmode=smart>. Acesso em: 15 mar. 2022.

Wu, F.-S., Hsu, C.-C., Lee, P.-C. & Su, H.-N. (2011). A systematic approach for integrated trend analysis-The case of etching. *Technological Forecasting and Social Change*, 78 (3), 386–407.

Xi, Y. & Xiang, Y. (2020). Patent analysis for graphene technology in ICT industry. In: *26th International Association for Management of Technology Conference, IAMOT 2017*, 1125–1141.