

UNIVERSIDADE FEDERAL DE SÃO CARLOS
DEPARTAMENTO DE CIÊNCIA DA INFORMAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

Ana Guelfi Erba

ESTRATÉGIAS DE ACESSO E MEDIAÇÃO PARA REPOSITÓRIO DE INFORMAÇÃO
TECNOLÓGICA INTEGRADA

São Carlos, SP

2026

Ana Guelfi Erba

ESTRATÉGIAS DE ACESSO E MEDIAÇÃO PARA REPOSITÓRIO DE INFORMAÇÃO
TECNOLÓGICA INTEGRADA

Dissertação de Mestrado apresentada ao
Programa de Pós-Graduação em Ciência da
Informação, vinculado ao Departamento de
Ciência da Informação da Universidade Federal
de São Carlos

Área de concentração: Conhecimento,
Tecnologia e Inovação
Linha de pesquisa 1: Conhecimento e
Informação para Inovação

Orientação: Luc Quoniam
Coorientação: Dênis Leonardo Zaniro

São Carlos, SP
2026

Erba, Ana Guelfi

Estratégias de acesso e mediação para repositório de
informação tecnológica integrada / Ana Guelfi Erba --
2026.
67f.

Dissertação (Mestrado) - Universidade Federal de São
Carlos, campus São Carlos, São Carlos

Orientador (a): Luc Quoniam

Banca Examinadora: Luc Quoniam, Wanda Aparecida
Machado Hoffmann, Fábio José Justo dos Santos

Bibliografia

1. Mediação da informação. 2. Ciência aberta. 3.
Informação tecnológica. I. Erba, Ana Guelfi. II. Título.

Ficha catalográfica desenvolvida pela Secretaria Geral de Informática
(SIn)

DADOS FORNECIDOS PELO AUTOR

Bibliotecário responsável: Arildo Martins - CRB/8 7180

Folha de Aprovação

Defesa de Dissertação Mestrado da candidata Ana Guelfi Erba, realizada em 24/04/2026.

Comissão Julgadora:

Prof. Dr. Luc Marie Quoniam (UFSCar)

Prof. Dra. Wanda Aparecida Machado Hoffmann (UFSCar)

Prof. Dr. Fábio José Justo dos Santos (IFSP)

Dedico este trabalho à minha mãe.

AGRADECIMENTOS

Agradeço aos membros da banca.

Agradeço aos professores e colegas do PPGCI.

Agradeço ao meu orientador, Prof. Dr. Luc Marie Quoniam, e ao meu coorientador, Dênis Leonardo Zaniro.

Agradeço ao IBICT pelo apoio financeiro na realização deste trabalho.

RESUMO

Este trabalho apresenta o desenvolvimento de um repositório integrado de efeitos científicos e tecnológicos, com o objetivo de mitigar a barreira de transferência da informação identificada em uma base de conhecimento originalmente estática. A pesquisa, de natureza aplicada e empírica, foi conduzida para viabilizar a conversão de um vasto estoque de soluções inventivas de problemas em um fluxo de conhecimento acessível e utilizável. Inicialmente, realizou-se uma avaliação diagnóstica da base de dados em planilha eletrônica, identificando lacunas críticas de interoperabilidade e acessibilidade que limitavam o uso estratégico dos dados, conforme preconizam as diretrizes internacionais para dados abertos. O percurso metodológico envolveu a estruturação lógica dos dados, adotando um tipo de dado específico para favorecer a representação coerente do conteúdo em múltiplos idiomas. Em seguida, foi desenvolvido um artefato de mediação, um sistema Web estruturado com base nos princípios de Usabilidade, visando otimizar a eficiência da interação humano-computador e da recuperação da informação. Para complementar as metodologias exatas de pesquisa relacional, a arquitetura do sistema incorporou recursos de Inteligência Artificial, através da técnica de Geração Aumentada por Recuperação. A solução tecnológica implementada atendeu aos objetivos propostos ao estabelecer múltiplas estratégias de acesso especializadas. O sistema possibilita buscas por diferentes perspectivas, conectando o conhecimento abstrato das soluções inventivas de problemas às classificações industriais e patentárias, o que agrega valor estratégico à informação tecnológica. A avaliação de aderência final do artefato evidenciou o alinhamento da arquitetura com os pilares de encontrabilidade, acessibilidade, interoperabilidade e reusabilidade. Conclui-se que o trabalho fornece um modelo empírico de mediação que reitera a relevância da área de Informação Científica na organização e na arquitetura de acesso do conhecimento para o desenvolvimento socioeconômico. A solução não apenas contribui para a ampliação do acesso ao recurso, mas também potencializa sua utilização como um ativo estratégico e reutilizável.

Palavras-chave: mediação da informação; ciência aberta; informação tecnológica.

ABSTRACT

This work presents the development of an integrated repository of scientific and technological effects, with the objective of mitigating the information transfer barrier identified in an originally static knowledge base. The research, of an applied and empirical nature, was conducted to enable the conversion of a vast stock of inventive problem-solving solutions into an accessible and usable knowledge flow. Initially, a diagnostic evaluation of the spreadsheet database was carried out, identifying critical gaps in interoperability and accessibility that limited the strategic use of the data, as advocated by international guidelines for open data. The methodological path involved the logical structuring of the data, adopting a specific data type to favor the coherent representation of the content in multiple languages. Subsequently, a mediation artifact was developed—a Web system structured based on Usability principles, aiming to optimize the efficiency of human-computer interaction and information retrieval. To complement the exact relational search methodologies, the system's architecture incorporated Artificial Intelligence features through the Retrieval-Augmented Generation (RAG) technique. The implemented technological solution met the proposed objectives by establishing multiple specialized access strategies. The system enables searches from different perspectives, connecting the abstract knowledge of inventive problem-solving solutions to industrial and patent classifications, which adds strategic value to technological information. The final adherence evaluation of the artifact evidenced the alignment of the architecture with the pillars of findability, accessibility, interoperability, and reusability. It is concluded that the work provides an empirical mediation model that reiterates the relevance of the Scientific Information area in the organization and architecture of knowledge access for socioeconomic development. The solution not only contributes to broadening access to the resource but also enhances its utilization as a strategic and reusable asset.

Keywords: information mediation; open science; technological information.

LISTA DE FIGURAS

Figura 1: Visão Geral das Tabelas do Artefato de Mediação.....	44
Figura 2: Interface de Busca.....	47
Figura 3: Busca por Função.....	48
Figura 4: Resultado da Busca por Função (Radioatividade)	49
Figura 5: Resultado da Busca por Função (Ablação por laser)	49
Figura 6: Resultado da Busca por Função - Idioma Alternativo	50
Figura 7: Mensagens de Erro para Buscas Incompletas	50
Figura 8: Busca por Parâmetro	51
Figura 9: Busca por Transformação	51
Figura 10: Busca por Efeito.....	52
Figura 11: Resultado da Busca por Efeito	53
Figura 12: Busca por CNAE.....	53
Figura 13: Resultado da Busca por CNAE.....	54
Figura 14: Opções de Busca da Perspectiva Patentária.....	54
Figura 15: Busca por Código IPC.....	55
Figura 16: Busca por Palavra-chave WIPO.....	56
Figura 17: Busca Semântica com IA	56
Figura 18: Fontes Técnicas do RAG	57

LISTA DE TABELAS

Tabela 1: Colunas da Base de efeitos TRIZ	38
---	----

LISTA DE ABREVIATURAS E SIGLAS

API	<i>Application Programming Interface</i> (Interface de Programação de Aplicações)
CC	Ciência da Computação
CC BY-SA 3.0	<i>Creative Commons Attribution-ShareAlike 3.0 Brazil</i>
CI	Ciência da Informação
CNAE	Classificação Nacional de Atividades Econômicas
CNPJ	Cadastro Nacional de Pessoa Jurídica
DDL	<i>Data Definition Language</i>
EPO	<i>European Patent Office</i> (Escritório Europeu de Patentes)
FAIR	<i>Findable, Accessible, Interoperable, Reusable</i>
FCPJ	Ficha Cadastral de Pessoa Jurídica
FID	Federação Internacional de Informação e Documentação
GIF	<i>Graphics Interchange Format</i>
GPU	<i>Graphics Processing Unit</i> (Unidade de Processamento Gráfico)
HTML	<i>Hypertext Markup Language</i>
HTTP	<i>Hypertext Transfer Protocol</i>
HTTPS	<i>Hypertext Transfer Protocol Secure</i>
IA	Inteligência Artificial
ID	Identificador (ou <i>Identifier</i>)
i18n	<i>Internationalization</i> (Internacionalização)
IBGE	Instituto Brasileiro de Geografia e Estatística
IHC	Interação Humano-Computador
IPC	<i>International Patent Classification</i> (Classificação Internacional de Patentes)
JSONB	<i>Binary JavaScript Object Notation</i>
LLM	<i>Large Language Model</i> (Grande Modelo de Linguagem)
PHP	<i>Hypertext Preprocessor</i>
RAG	<i>Retrieval-Augmented Generation</i> (Geração Aumentada por Recuperação)
RDF	<i>Resource Description Framework</i>
RI	Recuperação da Informação
SENAI	Serviço Nacional de Aprendizagem Industrial
SGBD	Sistema Gerenciador de Banco de Dados
SRI	Sistemas de Recuperação da Informação
TRIZ	Teoria da Solução Inventiva de Problemas

UFSCar Universidade Federal de São Carlos

URI *Uniform Resource Identifier*

URL *Uniform Resource Locator* (Localizador Padrão de Recursos)

WIPO *World Intellectual Property Organization* (Organização Mundial da Propriedade Intelectual)

XML *eXtensible Markup Language* (Linguagem de Marcação Extensível)

SUMÁRIO

1 INTRODUÇÃO	13
2 PRESSUPOSTOS TEÓRICOS/ FUNDAMENTAÇÃO TEÓRICA	16
2.1 CIÊNCIA DA INFORMAÇÃO, MEDIAÇÃO DA INFORMAÇÃO E O FLUXO INFORMACIONAL.....	16
2.2 INFORMAÇÃO TECNOLÓGICA E METODOLOGIA TRIZ	18
2.3 PATENTES E CLASSIFICAÇÃO IPC	20
2.4 PERSPECTIVA INDUSTRIAL E CLASSIFICAÇÃO CNAE	21
2.5 WIKIPÉDIA	23
2.6 BASE DE EFEITOS TRIZ MULTILÍNGUE	24
2.7 CIÊNCIA ABERTA	25
2.8 RECUPERAÇÃO DA INFORMAÇÃO	28
3 MATERIAIS E MÉTODOS	31
4 RESULTADOS	38
4.1 ANÁLISE DIAGNÓSTICA DA BASE DE EFEITOS TRIZ MULTILÍNGUE	38
4.2 AVALIAÇÃO DIAGNÓSTICA FAIR	41
4.3 ESTRUTURAÇÃO DAS TABELAS DO ARTEFATO DE MEDIAÇÃO.....	43
4.4 O ARTEFATO	46
6 CONSIDERAÇÕES FINAIS.....	60
REFERÊNCIAS	62

1 INTRODUÇÃO

A comunicação eficaz da informação entre os seres humanos, com base em suas necessidades informacionais e contextos sociais, é um dos focos da Ciência da Informação (CI) (Saracevic, 1996). A CI emergiu após a Segunda Guerra Mundial para buscar formas de lidar com o grande volume de informações da época, que foi intensificado até chegar na contemporaneidade. Atualmente, conforme Saracevic (1999), a CI mantém e aprimora suas buscas por soluções para problemas relacionados à informação, tendo como desafio o contexto multifacetado do ambiente *Web*. Saracevic (1999) pontua que, tendo como base os pilares do comportamento informacional humano e de técnicas de Recuperação da Informação (RI), a CI se articula para lidar com este desafio, com ênfase na RI como mecanismo para viabilizar o acesso e uso eficaz das informações.

Dentre os registros que a CI busca tornar acessíveis por meio da RI, destaca-se a informação tecnológica, devido ao seu caráter estratégico em processos de inovação (Castro, Jannuzzi e Mattos, 2007). Este conceito, segundo Alvares e Itaborahy (2021), é abrangente e engloba todo conhecimento de natureza econômica, mercadológica, gerencial ou social que, quando aplicado, resulta em aperfeiçoamento ou inovação. Ainda segundo os autores, o gerenciamento e recuperação eficaz deste tipo de informação é essencial no presente contexto mercadológico da sociedade, pois incentiva a concorrência e a transformação da informação tecnológica em insumos competitivos.

Nesse sentido, a produção e a disseminação da informação são entendidas como elementos fundamentais para gerar inovação, constituindo-se como recursos estratégicos da economia de um país. Segundo Castro, Jannuzzi e Mattos (2007), o processo de inovar depende inequivocamente da disponibilidade de informação, uma vez que é a informação produzida pela ciência e traduzida em tecnologia que alavanca a competitividade. Dessa forma, a informação tecnológica atua como um vetor essencial na transformação do conhecimento científico em resultado econômico, materializando a inovação apenas quando esta é efetivamente assimilada pelo mercado e gera valor (Castro, Jannuzzi e Mattos, 2007).

No contexto organizacional, é fundamental considerar a informação tecnológica advinda de patentes, que se trata de uma detalhada fonte sobre invenções realizadas ao redor do mundo (Barroso; Quoniam; Pacheco, 2009; WIPO, 2022). Entretanto, o volume de documentos patentários traz um desafio em relação à análise e recuperação da informação. Para contornar este desafio, na década de 1940, Genrich Altshuller propôs uma metodologia que, conforme Altshuller (1999), é derivada de uma análise manual de um conjunto amplo de patentes e

denominada como TRIZ - sigla russa que significa Teoria da Solução Inventiva de Problemas. Essa metodologia fornece um meio estruturado para solucionar problemas técnicos relacionados a conflitos ou contradições (Moehrle, 2005).

Aspectos da metodologia TRIZ têm sido adotados em áreas da engenharia por organizações em diversas partes do mundo (Li *et al.*, 2022). De acordo com Martin (2015) e Zaniro *et al.* (2024), bases de efeitos TRIZ têm sido implementadas para auxiliar a aplicação prática dessa metodologia. Normalmente, essas bases implementam um subconjunto da metodologia TRIZ, e ajudam a responder perguntas técnicas como, por exemplo, “Como mover líquido?”, “Como aumentar a temperatura?” etc. (Oxford Creativity, 2025).

Neste trabalho, foi escolhida como objeto de estudos a Base de efeitos TRIZ proposta por Zaniro e Quoniam (2024), que contém soluções para diferentes problemas científicos e de engenharia. A base proposta é um recurso pautado na Ciência Aberta, pois, conforme Silva e Silveira (2019), segue a premissa de ampliar o acesso a produções científicas e permite reprodutibilidade pela sociedade. Além disso, é um repositório aberto de informação tecnológica com capacidade multilíngue e de interoperabilidade, devido à integração dos dados com páginas da Wikipédia e à correlação com códigos da Classificação Internacional de Patentes (IPC) e da Classificação Nacional de Atividades Econômicas (CNAE). Dessa forma, ressalta-se que não se trata de uma tradução de dados, mas de um mapeamento semântico entre idiomas, que possibilita que grupos de diferentes regiões geográficas e culturas utilizem o conhecimento disponibilizado (Zaniro *et al.*, 2024).

A base de Zaniro e Quoniam (2024) encontra-se disponibilizada gratuitamente na *Web* por meio do repositório institucional da Universidade Federal de São Carlos (UFSCar), em formato de planilha eletrônica. A característica de sua disponibilidade promove acesso à comunidade científica, mas entende-se que o formato planilhado pode gerar alguns desafios para a recuperação da informação no que se refere à consulta e cruzamento da informação. Esta situação exemplifica a tensão, apontada por Barreto (1996), entre as duas funções essenciais dos agregados de informação: a função de produção de estoques e a função de transferência. O formato planilhado é uma decisão orientada pela racionalidade técnica, mas que pode criar barreiras para a transferência, afetando assim o destino da geração de conhecimento.

Na perspectiva de recuperação da informação, é possível questionar “Como reestruturar e mediar, pela ótica da Ciência da Informação, os dados da Base de efeitos TRIZ Multilíngue para otimizar a recuperação de informação e maximizar seu acesso pela comunidade científica?”. Para responder a essa questão, faz-se necessário, preliminarmente, compreender a natureza do objeto a ser representado, ou seja, os recursos informacionais. Conforme elucida

Araújo (2010), a Ciência da Informação não se pauta em um conceito único e estático de informação, mas sim na coexistência e complementaridade de dimensões físicas, cognitivas e sociais. Sob essa ótica plural, o repositório TRIZ estrutura dados provenientes de perspectivas distintas — abrangendo desde a informação científica e enciclopédica até a inventiva —, configurando-se como parte de um sistema complexo e em evolução.

Assim, a representação proposta buscará compreender uma faceta funcional dessa complexidade, ciente de que a heterogeneidade das fontes exige uma abordagem integradora. Desse modo, justifica-se a necessidade de estudar e aplicar métodos de recuperação da informação que ampliem o acesso à base e permitam maior utilização de sua informação tecnológica no contexto de inovação organizacional, além de explorar como disponibilizar dados abertos para facilitar usos reais na perspectiva mercadológica.

Para materializar essa mediação, o desenvolvimento de um sistema Web se apresenta como uma alternativa acessível e prática. Um sistema Web pode ser definido como uma aplicação de software baseada em arquitetura cliente-servidor, que utiliza tecnologias da internet para fornecer conteúdo e funcionalidades interativas aos usuários por meio de navegadores (Pressman e Maxim, 2016). Essa abordagem viabiliza a implementação de estratégias de busca tradicionais, estruturando a recuperação da informação de forma clara e objetiva. Adicionalmente, uma vez estabelecida, essa arquitetura pode se estender para incorporar recursos de Inteligência Artificial (IA) – área voltada ao projeto e à construção de agentes que recebem percepções do ambiente e executam ações (Russell e Norvig, 1995).

Portanto, o objetivo geral deste trabalho constitui-se em propor um modelo de recuperação da informação para a Base de efeitos TRIZ Multilíngue, com a intenção de solucionar as barreiras de transferência da informação impostas em seu formato original. Para atingi-lo, propõem-se como objetivos específicos:

- Realizar uma análise da estrutura informacional da Base de efeitos TRIZ proposta por Zaniro e Quoniam, 2024 para identificar suas características e relações.
- Explorar outras bases de efeitos TRIZ disponíveis em meio digital para identificar padrões e possibilidades de melhorias pautadas em experiências consistentes.
- Propor uma modelagem estruturada para organizar a informação da Base de efeitos TRIZ de modo a preservar suas características multilíngues.
- Desenvolver um artefato para mediação capaz de intermediar a comunidade científica e suas ações de recuperação da informação.
- Realizar uma validação do artefato de mediação proposto, focada na sua capacidade de otimizar a recuperação da informação.

Para alcançar os objetivos propostos, este trabalho está estruturado em cinco seções. Esta seção contextualiza o problema de pesquisa, delimita a justificativa e estabelece os objetivos geral e específicos. A segunda seção, Fundamentação Teórica, aprofunda os conceitos centrais que embasam o estudo, com ênfase na Ciência da Informação, na Recuperação da Informação, na Informação Tecnológica, na metodologia TRIZ e nos preceitos da Ciência Aberta. A terceira seção, Materiais e Métodos, detalha o percurso metodológico adotado, incluindo a análise da base de efeitos original, a modelagem dos dados e as etapas de desenvolvimento do artefato de mediação. A quarta seção, Resultados, apresenta o modelo de recuperação da informação implementado e discute suas potencialidades para a solução das barreiras de transferência identificadas. Finalmente, a quinta seção apresenta as Considerações Finais, sintetizando as contribuições do trabalho e indicando possibilidades para pesquisas futuras.

2 PRESSUPOSTOS TEÓRICOS/ FUNDAMENTAÇÃO TEÓRICA

Esta seção apresenta a compreensão da literatura sobre os conceitos de Ciência da Informação, recuperação da informação, informação tecnológica, metodologia TRIZ, Ciência Aberta, que estão atrelados à pesquisa e seus objetivos.

2.1 CIÊNCIA DA INFORMAÇÃO, MEDIAÇÃO DA INFORMAÇÃO E O FLUXO INFORMACIONAL

O conceito de informação, fundamental para a Ciência da Informação, revela-se complexo e multifacetado, marcado por sua evolução ao longo da história (Araújo, 2010). Originalmente, o termo latino *informatio* estava ligado a um sentido ontológico, conectado aos conceitos gregos de *eidos* (ideia) e *morphé* (forma), significando o ato de "dar forma a algo" (Capurro, 2009, p. 125). Capurro (2009) destaca que os sistemas técnicos de transmissão da informação surgidos a partir do século XX, contribuíram para a substituição dos processos artificiais clássicos pela tecnologia digital, a partir da qual a tela assume o lugar do livro como metáfora central.

Nesse contexto comunicacional, a Ciência da Informação emerge, segundo Saracevic (1975), como a disciplina focada nos problemas decorrentes da comunicação do conhecimento. Sua finalidade primordial é assegurar a comunicação eficaz do conhecimento, dedicando-se, portanto, ao estudo do fluxo de informação. Desta forma, a CI se ocupa em otimizar esse fluxo, gerenciando a distinção crucial entre o que é meramente transmitido e o que é efetivamente

relevante para o usuário, um desafio central na era de mensageiros digitais previamente apontada.

Para que o fluxo de informação relevante seja efetivado, a Ciência da Informação aplica o conceito de mediação da informação. Esta mediação é entendida como o mecanismo central da área, definindo-se como toda ação de interferência (Silva, 2015, p. 103). De forma crucial para o contexto digital, essa intervenção não é apenas interpessoal, mas também sistêmica, ocorrendo por meio da mediação técnica da informação (Silva, 2015, p. 105). Esta dimensão técnica refere-se especificamente às ações de organização e representação da informação executadas nos sistemas, desde a elaboração de catálogos até a interação por redes sociais, conectando assim a prática da CI aos mensageiros digitais contemporâneos para satisfazer as necessidades dos usuários.

A mediação técnica, sobretudo em ambientes virtuais, expõe duas características estruturais da Ciência da Informação. A primeira é sua forte vocação interdisciplinar (Saracevic, 1995). A segunda é sua dependência intrínseca da tecnologia: Saracevic (1995) afirma que a CI está "inexoravelmente conectada à tecnologia da informação", o que define sua evolução. O próprio objetivo do campo pressupõe o uso intensivo dessas tecnologias. Nesse contexto, a recuperação da informação consolidou-se como uma das principais atividades de pesquisa da área, tornando-se, por consequência, o maior vetor de suas relações interdisciplinares (Saracevic, 1995).

Historicamente, a expressão "Recuperação da Informação" foi cunhada por Calvin Mooers em 1951. Embora proviesse originalmente da Ciência da Computação, a recuperação da informação passou a fazer parte integrante da Ciência da Informação em meados da década de 1960 (Saracevic, 1999). Neste contexto, a RI, sendo uma das atividades centrais da CI, constitui o principal ponto de intersecção com a Ciência da Computação (CC), uma vez que ambas as áreas possuem a informação como objeto comum (Cafezeiro; Costa; Kubrusly, 2016). Esta relação é recíproca: a CC oferece os mecanismos automatizados para a manipulação e armazenamento que viabilizam a RI em larga escala, enquanto a CI estabelece os desafios de modelagem da informação e os contextos de uso que direcionam a arquitetura desses sistemas (Cafezeiro; Costa; Kubrusly, 2016).

A partir desta intersecção, enfatiza-se a oportunidade de automatizar a complexa tarefa da organização e representação da informação no âmbito da CI, fundamental para a RI. Segundo De Albuquerque (2015), este processo envolve identificar semelhanças para agrupar informações, permitindo que sejam acessadas de forma precisa. Embora este conceito seja vital para a CI, os métodos de organização podem se tornar obsoletos rapidamente, seja pela

velocidade com que novas informações são adicionadas ou conforme as necessidades sociais se alteram (De Albuquerque, 2015).

O desafio da organização da informação requer mais mecanismos do que somente armazenar dados. Essa dinâmica remete à dualidade explorada por Buckland (1991), na qual a 'informação como coisa' (o artefato ou dado armazenado) precisa ser ativada por métodos eficientes de mediação técnica para se tornar 'informação como processo', viabilizando seu uso e transferência ao usuário. Um "estoque" de informação, por mais valioso que seja, torna-se funcionalmente obsoleto se sua organização e os métodos de recuperação — a mediação técnica — não forem eficientes para promover o "fluxo" de conhecimento.

Este cenário pode ser crítico em domínios especializados, como o da informação tecnológica. Nesse domínio, de acordo com Alvares e Itaborahy (2021), a complexidade advém especialmente de sua natureza, cuja aplicação precisa mobilizar perspectivas e agentes diferentes para a resolução de problemas no contexto organizacional. Esta característica exige soluções de organização e representação que superem as barreiras do armazenamento. Diante disso, torna-se essencial analisar as especificidades deste tipo de informação e dos sistemas que buscam estruturá-la, como a metodologia TRIZ.

2.2 INFORMAÇÃO TECNOLÓGICA E METODOLOGIA TRIZ

A informação tecnológica é um termo abrangente que consolida as singularidades das informações técnicas, econômicas, organizacionais e de engenharia (Alvares e Itaborahy, 2021). Em complemento, a Federação Internacional de Informação e Documentação (FID) define, a informação tecnológica diz respeito a “todo conhecimento de natureza técnica, econômica, mercadológica etc. que, por sua aplicação, favoreça o progresso na forma de aperfeiçoamento e inovação” (Araújo, 1993, p.228). Por ser um insumo crítico para a inovação, ela engloba todo o conhecimento necessário à organização, seja ele interno ou externo, de livre acesso ou proprietário (Alvares e Itaborahy, 2021). Um aspecto importante é que esse conhecimento pode fluir por canais formais, como bancos de dados, e por canais que customizam e agregam valor ao usuário final (Alvares e Itaborahy, 2021).

O desenvolvimento tecnológico é, portanto, indissociável da obtenção e, sobretudo, do uso estratégico da informação tecnológica. O Serviço Nacional de Aprendizagem Industrial (2002) corrobora essa visão ao defini-la como todo conhecimento efetivamente aplicado ao desenvolvimento e à produção de tecnologias ou serviços, tendo como finalidade central a melhoria contínua dos setores produtivos. Ela é valorizada por sua aplicação direta na solução de problemas práticos e na geração de valor econômico. Por essa razão, este tipo de informação

frequentemente constitui um ativo proprietário ou confidencial, essencial para a competitividade organizacional, cujo fluxo, como observado, muitas vezes ocorre por canais informais que agregam valor direto ao usuário.

A metodologia TRIZ surge precisamente como uma resposta ao desafio de extrair valor da informação tecnológica. A TRIZ diferenciou-se de métodos anteriores, como o *brainstorming*, ao focar sua análise exatamente nos registros do produto criativo das áreas técnicas: as patentes (Carvalho, 2017). Altshuller analisou essa vasta base de informação tecnológica buscando regularidades que pudessem ser usadas para a solução de problemas futuros. O resultado, segundo Altshuller (1999), é uma metodologia definida como heurística, orientada ao ser humano e baseada em conhecimento, para a resolução de problemas inventivos. A TRIZ foca, especificamente, problemas inventivos, notáveis por exigirem a resolução de uma contradição.

Conforme Carvalho (2017), sua estratégia central é a abstração (mapear um problema específico para um problema genérico), seguida pela resolução (usar ferramentas como os Efeitos para encontrar uma solução) e, finalmente, a particularização (adaptar a solução genérica ao problema específico). Para operacionalizar essa busca por soluções, as bases de efeitos TRIZ desempenham um papel fundamental. Elas compõem uma ampla gama de ferramentas da metodologia cujo objetivo é aproximar o conhecimento teórico do contexto organizacional.

A aplicação prática da TRIZ, especialmente o uso dos Efeitos oriundos das ciências e engenharias, é o que permite a recuperação efetiva da informação tecnológica. Essa recuperação é frequentemente estruturada por um modelo funcional que define a necessidade de informação do usuário. Este modelo, conforme exemplificado na introdução, baseia-se na trinca tarefa-alvo-efeito. A tarefa (ou função) representa a ação desejada (como "mover" ou "aumentar"), enquanto o alvo é o objeto sobre o qual a ação incide (como "líquido" ou "temperatura"). Já o efeito, ou conjunto de efeitos, representa sugestões de princípios científicos e de engenharia, ou ainda técnicas que, naturalmente ou de maneira aplicada, podem levar a uma solução envolvendo a tarefa e o alvo. Este modelo funcional é a ponte entre a metodologia TRIZ e os sistemas de recuperação da informação.

Este modelo funcional (tarefa-alvo-efeito) é, portanto, a chave para transformar a metodologia TRIZ em um estoque de informação tecnológica estruturada, que é um recurso dinâmico pronto para o uso e a assimilação, conforme a perspectiva de Buckland (1991). Diversas bases de dados buscam implementar esse modelo, e uma dessas bases é a proposta por Zaniro e Quoniam (2024), objeto desta pesquisa. No entanto, a criação de um "estoque" digital,

por si só, não garante sua "transferência" ou uso efetivo. A base de Zaniro e Quoniam (2024) avança nessa questão ao ser concebida sob os preceitos da Ciência Aberta, disponibilizada publicamente no repositório da UFSCar. Contudo, para que o acesso democrático se concretize, a disponibilidade não é suficiente, mas é preciso que os dados sigam princípios que garantam sua real localização, acessibilidade e reuso.

2.3 PATENTES E CLASSIFICAÇÃO IPC

Aprofundando a discussão sobre os canais formais de informação tecnológica introduzida na seção anterior, é fundamental analisar a fonte primária que deu origem à própria metodologia TRIZ: o sistema de patentes. Se a eficácia da transferência da informação depende da organização de seu "estoque" (Barreto, 1996), os documentos de patentes representam um amplo repositório estruturado de conhecimento técnico e científico. Eles materializam a informação tecnológica descrita por Alvares e Itaborahy (2021), servindo não apenas como instrumentos de proteção à propriedade intelectual, mas como vetores essenciais de disseminação do estado da técnica (Barros *et al.*, 2014).

Segundo a Organização Mundial da Propriedade Intelectual (WIPO, 2022), as patentes contêm informações técnicas que muitas vezes não são divulgadas em nenhum outro meio, tornando-as insumos importantes para fomentar a inovação. Nesse sentido, Paranaguá e Reis (2009) esclarecem que, embora as patentes de invenção visem assegurar a titularidade e a exclusividade de exploração de uma inovação, elas atuam fundamentalmente como um instrumento de divulgação científica. Para os autores, a descrição detalhada do invento é a "contrapartida social" desse direito de exclusividade: ao tornar públicas as informações descritivas, o sistema impede o surgimento de segredos industriais e garante que o conhecimento tecnológico seja revelado à sociedade para impulsionar o desenvolvimento (Paranaguá; Reis, 2009).

Contudo, para que esse vasto acervo de conhecimento seja efetivamente acessível, é necessário um sistema de organização robusto. É nesse contexto que se insere a Classificação Internacional de Patentes (IPC, ou CIP em português), um sistema hierárquico administrado pela Organização Mundial da Propriedade Intelectual (WIPO, ou OMPI em português) que possibilita a organização da informação tecnológica, seja pelo tipo de técnica utilizada ou pela função do produto tecnológico. De acordo com Garcia e Chacon (2008), a IPC funciona como uma ferramenta de busca que utiliza uma listagem de itens importantes para acessar a informação contida nos documentos, provendo uma recuperação precisa da tecnologia e facilitando o acesso à matéria técnica.

Para auxiliar a navegação nessa estrutura, a própria organização da IPC disponibiliza um índice oficial conhecido como *Catchwords*. Segundo a WIPO (2024), esse guia atua como um vocabulário alfabético de palavras-chave que direciona o usuário aos códigos hierárquicos correspondentes. O recurso é apresentado nos idiomas inglês e francês, orientando a localização de conceitos técnicos dentro da classificação sem a necessidade de conhecer previamente o código exato da patente.

Para a operacionalização da recuperação da informação tecnológica organizada pela IPC, utilizam-se sistemas de busca especializados que agregam esses dados em escala global. Nesse cenário, o Espacenet, mantido pelo Escritório Europeu de Patentes (EPO), destaca-se como uma das ferramentas mais robustas e acessíveis, contendo a maior base de dados de patentes do mundo. Segundo o *European Patent Office* (2025), a plataforma disponibiliza acesso gratuito a mais de 160 milhões de documentos de patentes. Essa abrangência reforça a análise de Pires, Ribeiro e Quintella (2020), que ressaltam a relevância do sistema para a verificação do estado da técnica. Os autores concluem que, para buscas utilizando códigos de classificações internacionais (como a IPC), o *Espacenet* oferece resultados de alta precisão, essenciais para monitorar inovações, concorrentes e identificar parceiros de transferência de tecnologia, democratizando o acesso a informações que antes eram restritas a sistemas proprietários de alto custo.

Dessa forma, a integração da Classificação Internacional de Patentes na estrutura da Base de efeitos TRIZ Multilíngue não constitui apenas um recurso de organização, mas uma estratégia fundamental de mediação da informação. Ao vincular os efeitos científicos — que representam a essência abstrata da resolução de problemas na TRIZ — aos códigos hierárquicos da IPC, o repositório restabelece o vínculo da metodologia com sua fonte primária: os documentos de patentes. Essa conexão permite transformar o 'estoque' de conhecimento técnico em um fluxo navegável, possibilitando que o usuário transite do princípio científico teórico até suas aplicações tecnológicas concretas, contextualizando a invenção dentro dos domínios formalizados da propriedade industrial.

2.4 PERSPECTIVA INDUSTRIAL E CLASSIFICAÇÃO CNAE

A organização do conhecimento tecnológico, estruturada pelo sistema de patentes e pela IPC, alcança sua finalidade social e econômica quando transposta para o cenário produtivo. A transformação da invenção em inovação ocorre, primordialmente, no ambiente industrial e de mercado, local onde o conhecimento técnico codificado nos documentos de patentes se converte em valor tangível e competitividade.

Nesse contexto, a interação entre os detentores do conhecimento e os centros produtivos apresenta-se como fator fundamental para o desenvolvimento tecnológico. Oliveira Junior e Almeida (2019) ressaltam que a propriedade intelectual necessita ser analisada sob uma perspectiva estratégica de mercado, e não apenas como um mecanismo burocrático de proteção. Para os autores, a transferência de tecnologia atua como um instrumento para o incremento econômico, permitindo que os resultados de pesquisas científicas e tecnológicas sejam efetivamente incorporados pelas empresas.

Contudo, observa-se que a existência de um ativo protegido por patente não implica, automaticamente, sua absorção pelo mercado. Oliveira Junior e Almeida (2019) apontam a existência de desafios no alinhamento entre as tecnologias desenvolvidas nos centros de pesquisa e as demandas reais da indústria. A gestão da propriedade intelectual, portanto, demanda um olhar voltado para as necessidades dos setores produtivos, buscando superar a lacuna entre a oferta tecnológica acadêmica ou inventiva e a capacidade de industrialização e comercialização desses ativos (Oliveira Junior; Almeida, 2019).

Para operacionalizar essa aproximação com o mercado e permitir a identificação precisa dos setores produtivos, torna-se necessário o uso de um padrão que represente a realidade econômica nacional. Nesse sentido, a Classificação Nacional de Atividades Econômicas (CNAE) atua como o instrumento fundamental de organização e padronização.

Sob a gestão técnica do Instituto Brasileiro de Geografia e Estatística (IBGE), a CNAE constitui o padrão nacional para os códigos de atividade econômica e critérios de enquadramento utilizados pelos diversos órgãos da Administração Tributária (IBGE, 2025). Sua aplicação abrange todos os agentes econômicos engajados na produção de bens e serviços, compreendendo desde empresas privadas e públicas e estabelecimentos agrícolas até instituições sem fins lucrativos e agentes autônomos (IBGE, 2022).

Para garantir a precisão no enquadramento dos setores econômicos, a base de efeitos TRIZ adota a CNAE versão 2.3, atualização da taxonomia que refina o detalhamento das atividades para fins de registro público. Segundo o IBGE (2025), essa classificação é estruturada hierarquicamente em cinco níveis de desagregação: seção, divisão, grupo, classe e subclasse. Este último nível, a subclasse, composta por sete dígitos, assume papel central na modelagem do artefato, pois é definido especificamente para atender às demandas da Administração Pública e constitui o código padrão do Cadastro Nacional de Pessoa Jurídica (CNPJ), viabilizando o vínculo direto entre os efeitos científicos e a realidade empresarial brasileira.

Dessa forma, a incorporação da CNAE ao modelo de recuperação da informação proposto neste trabalho visa traduzir a "perspectiva de mercado" em dados estruturados. Ao conectar os efeitos TRIZ e as patentes (IPC) aos códigos CNAE, estabelece-se um vínculo direto com a base formal da economia brasileira, permitindo que a tecnologia seja localizada não apenas por sua função técnica, mas pelo setor industrial ou de serviços apto a implementá-la.

2.5 WIKIPÉDIA

A organização do conhecimento não se restringe aos padrões industriais formalizados como a CNAE, estendendo-se também a ambientes digitais de construção coletiva que ampliam o acesso à informação para além das fronteiras corporativas. Nesse cenário, a Wikipédia consolida-se como uma enciclopédia *online* baseada no modelo *wiki*, caracterizada pela colaboração comunitária, na qual usuários podem editar e manter conteúdos organizados em páginas e espaços de discussão (Piscopo *et al.*, 2017).

Diferentemente de bases de dados puramente estruturadas, a plataforma privilegia o uso da linguagem natural para tornar o conteúdo acessível à leitura humana, regendo-se por políticas de verificabilidade que exigem que as informações inseridas nos artigos sejam sustentadas por fontes externas confiáveis, publicadas e com autoria identificável, visando garantir a credibilidade do conteúdo disseminado (Piscopo *et al.*, 2017).

A relevância da plataforma estende-se para além do acúmulo de saber, impactando diretamente resultados econômicos no mundo real. Hinnosaar *et al.* (2019) demonstram que o conteúdo gerado por usuários exerce uma influência causal sobre o comportamento do consumidor. Segundo os autores, o incremento na quantidade e qualidade das informações disponíveis nas páginas da Wikipédia — como a adição de textos e imagens — correlaciona-se positivamente com o aumento da demanda por serviços associados àquele conteúdo, evidenciando que a plataforma atua como um mecanismo eficiente para mitigar assimetrias de informação e orientar escolhas individuais.

Nesse cenário, a Wikipédia apresenta-se como uma ferramenta estratégica de "presença online". Diferentemente de bases de dados estáticas, seu conteúdo dinâmico e multilíngue permite superar barreiras de acesso à informação, transformando o conhecimento digital em valor econômico tangível. A melhora na representação de um tema na plataforma tende a gerar retornos significativos, sugerindo que o conteúdo em bens públicos digitais possui um valor intrínseco para o desenvolvimento econômico e para a redução de desigualdades informacionais.

A perspectiva multilíngue da Wikipédia transcende a mera tradução de conteúdos, estabelecendo-se como uma infraestrutura vital para a preservação cultural e o acesso à informação em contextos em que recursos digitais são escassos. Krisnawati e Mahastama (2019) destacam que, para línguas classificadas como "com poucos recursos" (*under-resourced*) ou críticas - a exemplo do indonésio e do javanês, a plataforma atua como uma das fontes mais proeminentes de documentos digitais, preenchendo lacunas deixadas pela falta de acervos tradicionais digitalizados.

Mais do que replicar dados, as autoras demonstram que a construção colaborativa nesses idiomas captura a realidade cultural viva e seus vieses intrínsecos; no caso da Wikipédia javanesa, por exemplo, os artigos refletem uma diversidade linguística autêntica ao incorporar variados dialetos e níveis de fala (como *Ngoko* e *Krama*), contrastando com a homogeneidade estilística encontrada em revistas tradicionais. Dessa forma, a enciclopédia não apenas mitiga assimetrias informacionais, mas viabiliza o desenvolvimento de tecnologias locais, servindo de base para pesquisas de processamento de linguagem natural que dependem dessa riqueza de dados culturalmente situados.

Dessa forma, a integração da Wikipédia à arquitetura da Base de efeitos TRIZ Multilíngue transcende a função de um glossário auxiliar, constituindo-se como uma estratégia de enriquecimento semântico e cultural. Ao vincular os efeitos TRIZ aos artigos da enciclopédia, o artefato proposto apropria-se da capilaridade linguística e da validação comunitária descritas, garantindo que a definição dos princípios científicos não fique restrita a uma tradução literal, mas se beneficie da diversidade de contextos e da atualização contínua oferecida pela plataforma. Essa conexão permite que a base de dados supere as barreiras de compreensão técnica, transformando o conceito abstrato da TRIZ em um conhecimento contextualizado e acessível em múltiplos idiomas, alinhando-se aos objetivos de democratização e transferência da informação tecnológica.

2.6 BASE DE EFEITOS TRIZ MULTILÍNGUE

Conforme descrito, a base de dados de efeitos TRIZ escolhida como objeto de estudo para este trabalho possui característica multilíngue. Esta característica foi alcançada através da integração dos dados de efeitos TRIZ com artigos da Wikipédia (Hinnosaar *et al.*, 2023), considerando atualmente quatro idiomas: inglês, português, espanhol e francês. Para esse propósito, foi utilizada a API (em inglês, *Application Programming Interface* - *Interface de Programação de Aplicações*) fornecida pela Wikipédia.

Essa base de efeitos TRIZ, composta por efeitos presentes nas bases *Aulive* e *Oxford Effects Database*, estabelece uma ligação entre códigos CNAE e códigos IPC de modo que processos inventivos, patentes e setores econômicos brasileiros possam ser correlacionados (Zaniro e Quoniam, 2024). Dessa forma, essa base de dados representa uma aproximação entre soluções tecnológicas e setores industriais, criando caminhos para a aplicação prática da informação tecnológica.

Como em outras bases de efeitos TRIZ, os modelos de solução são especificados basicamente a partir da relação entre tarefa (*task*), alvo (*target*) e efeito (*effect*). A tarefa é uma ação a ser realizada para solucionar um dado problema. O alvo identifica o objeto ou matéria à qual a tarefa será aplicada. O efeito é o princípio científico ou fenômeno físico a ser obtido na resolução do problema. Na pergunta “Como transformar energia mecânica em química?”, “transformar energia mecânica” é a tarefa, “energia química” é o alvo e, para essa relação, existe um conjunto de efeitos associados - ou possibilidades de solução -, como uso de acumulador, soldagem por explosão, soldagem por fricção etc.

Na base de efeitos TRIZ de Zaniro e Quoniam (2024), há elementos complementares à relação entre tarefa, alvo e efeito. A estrutura inclui o modo (*mode*) de aplicação da tarefa - pode ser uma função, parâmetro ou transformação e o tipo do efeito - se é um efeito puro ou aplicado a partir de algum princípio científico. Na pergunta “Como transformar energia mecânica em química”, o modo refere-se à uma transformação, e a solução envolve efeitos puros e aplicados. Cada efeito é identificado por um nome (*name*), pode ser complementado por uma descrição (*description*), que detalha o contexto, e por um exemplo (*example*) que apresenta aplicações práticas.

Cada entrada da base de dados TRIZ também pode conter observações extras (*note*) e normalmente contém *links* apontando para as páginas da Wikipédia referentes ao conceito, nos quatro idiomas contemplados - inglês (*EN link*), português (*PT link*), francês (*FR link*) e espanhol (*ES link*). Verifica-se, assim, que os elementos complementares permitem enriquecer a compreensão e a aplicabilidade dos efeitos da base.

2.7 CIÊNCIA ABERTA

Tradicionalmente, a ciência enfrenta desafios como o acesso desigual a recursos e concentração de informações em comunidades científicas específicas. Um novo paradigma colaborativo tem ganhado força para ampliar a comunicação e o compartilhamento científico para viabilizar o acesso às produções da ciência. A transparência da pesquisa científica em todas as suas fases é incentivada pelo movimento chamado Ciência Aberta (Silva e Silveira, 2019).

De acordo com Silva e Silveira (2019), a Ciência Aberta também promove clareza e divulgação de métodos, resultados e dados científicos, com o objetivo de ampliar a reprodutibilidade científica e disponibilidade para todos os níveis da sociedade.

A Ciência Aberta é composta por pilares que são definidos de diferentes formas, de acordo com a literatura estudada. Conforme Araújo *et al* (2021), os pilares variam em suas terminologias, mas são pautados nos princípios de Ciência Aberta. Por meio de uma taxonomia brasileira proposta por Araújo *et al* (2021), a perspectiva dos especialistas brasileiros propõe como pilares fundamentais acesso aberto, dados abertos, pesquisa reprodutível, avaliação aberta, ciência cidadã e educação aberta. Essa proposta é baseada em facetas filosóficas, científicas e sociais, porque são pautadas em ética, integridade, inovação, reprodutibilidade, reuso, colaboração em redes, ciência cidadã e democratização do conhecimento.

Segundo Araújo *et al* (2021), acesso aberto diz respeito à disponibilização de resultados científicos sem barreiras financeiras ou legais; dados abertos se referem ao compartilhamento dos dados de pesquisas científicas de modo que permita o reuso; pesquisa reprodutível é pautada na transparência das práticas e metodologias científicas para permitir reprodutibilidade; avaliação aberta se refere a revisão por pares aberta e a possibilidade de medir impacto científico; ciência cidadã é a inclusão de não cientistas nas pesquisas científicas; educação aberta se refere à disseminação de recursos educacionais para aumentar o acesso ao conhecimento.

Em 2014, um consenso sobre a necessidade de uma infraestrutura global para dados abertos surgiu por meio de manifestações referentes aos princípios FAIR, na conferência Jointly designing a data FAIRPORT (Henning *et al*, 2019). Os princípios FAIR (*Findable, Accessible, Interoperable, Reusable*) são uma proposta de práticas e princípios para incentivar e protocolar o acesso aberto, de modo a colaborar com a interoperabilidade e reutilização de dados científicos com maior facilidade.

Wilkinson *et al.* (2016) propuseram o FAIR *Guiding Principles*, com o objetivo de estabelecer as práticas recomendadas para cada componente do acrônimo FAIR. Este guia contribui para o pilar de dados abertos da Ciência Aberta, uma vez que direciona os pesquisadores para ações padronizadas que amplificam o potencial dos dados abertos. As definições do guia são apresentadas a seguir.

Conforme Wilkinson *et al.* (2016), para que os dados sejam encontráveis (*Findable*), é necessário que tenham um identificador único que não pode ser alterado. Além disso, os dados devem estar acompanhados por informações e metadados detalhados sobre seu conteúdo e contexto. O identificador único precisa estar explícito nos dados para propiciar clara relação

entre os dados e suas descrições. Os dados devem ser registrados ou indexados em recursos que permitam sua recuperação e pesquisa de forma acessível.

No que se refere a dados acessíveis (*Accessible*), Wilkinson *et al.* (2016) propõem que os dados sejam recuperáveis por seu identificador único através de um protocolo de comunicação padronizado, de preferência um protocolo aberto, gratuito e universalmente aceito, que permita autenticação se houver necessidade. Também é importante que os dados sejam acessíveis mesmo se estiverem indisponíveis (removidos, por exemplo), para permitir acesso ao seu contexto ou histórico.

Quanto à interoperabilidade dos dados (*Interoperable*), Wilkinson *et al.* (2016) recomendam que os dados utilizem uma linguagem formal e acessível para sua representação, como padrões que sejam reconhecidos por diferentes sistemas, e que também sejam baseados em vocabulários padronizados. Além disso, os dados devem possuir referências qualificadas para outros dados, estabelecendo conexões relevantes e ampliando o contexto do conteúdo.

Para que os dados sejam reutilizáveis (*Reusable*), Wilkinson *et al.* (2016) indicam que sejam descritos com atributos plurais, precisos e relevantes, disponibilizados sob uma licença de uso explícita, acessível e sem ambiguidades. Também é importante que esses dados possuam informações detalhadas sobre sua origem ou obtenção e que atendam aos padrões comunitários da comunidade científica da área à qual o dado pertence.

Nesse contexto de licenciamento explícito, as licenças *Creative Commons* (CC) surgem como instrumentos jurídicos padronizados que flexibilizam o direito autoral tradicional para facilitar o reuso. Dentre as modalidades disponíveis, destaca-se a *Attribution-ShareAlike 3.0 Brazil* (CC BY-SA 3.0), fundamental para projetos colaborativos. De acordo com a *Creative Commons* Brasil (2025), esta licença permite que terceiros copiem, distribuam, exibam e executem a obra, bem como criem obras derivadas, sob duas condições: deve-se dar o crédito ao autor original e as novas criações devem ser licenciadas sob termos idênticos. Essa cláusula de reciprocidade é importante para a Ciência Aberta, pois assegura que as melhorias e adaptações feitas sobre um conjunto de dados permaneçam abertas para toda a comunidade.

É importante enfatizar que, com aumento exponencial de dados disponibilizados, principalmente dados científicos potencializados pela Ciência Aberta, é necessário que recursos de recuperação e acesso a dados sejam confiáveis, acessíveis e possam ter uma gestão adequada dos dados, para que a ética e reprodutibilidade sejam viabilizadas (Lin *et al.*, 2020). Neste sentido, os princípios FAIR representam um movimento importante na Ciência Aberta.

Os padrões propostos por FAIR permitem contextualizar dados e ampliar sua utilidade, além de aumentarem a eficiência na recuperação de informação (Henning *et al.*, 2018).

Conforme Henning *et al.* (2018), a adoção destes princípios nos dados científicos resultantes de pesquisas deve levar em consideração os âmbitos técnicos e culturais inerentes às pesquisas, porque sua aplicação tem impacto direto no engajamento social e na contribuição econômica por meio destes dados abertos, fazendo com que os princípios FAIR estabeleçam total consonância com a Ciência Aberta.

Os princípios FAIR, alinhados à Ciência Aberta, são essenciais para promover a transparência, colaboração e divulgação de dados científicos no paradigma de dados abertos, uma vez que têm potencial para gerir e compartilhar dados de forma sistemática, localizável, acessível, interoperável e reutilizável. Além disso, dada a importância de democratizar o acesso ao conhecimento científico, a base de efeitos TRIZ multilíngue elaborada por Zaniro e Quoniam (2024) é um recurso fundamental disponibilizado gratuitamente na *Web*. Assim, entende-se a possibilidade de explorar como os princípios FAIR podem ser aplicados à base de efeitos TRIZ multilíngue, para não somente facilitar o acesso e compreensão da base por parte da comunidade científica, mas também contribuir para o avanço das práticas de Ciência Aberta.

2.8 RECUPERAÇÃO DA INFORMAÇÃO

A Recuperação da Informação (RI) é frequentemente descrita como a vertente tecnológica da Ciência da Informação, sendo um resultado direto da relação interdisciplinar desta com a Ciência da Computação (Saracevic, 1999). Embora o termo seja usado de forma ampla, seu processo central consiste em identificar, no conjunto de documentos (corpus) de um sistema, quais atendem à necessidade de informação do usuário (Ferneda, 2003). Ferneda (2003) destaca que os sistemas não recuperam a informação em si, mas sim documentos ou referências cujo conteúdo poderá ser relevante", tratando-se de uma informação potencial que só se concretiza na análise do usuário.

O alicerce da RI é, por definição, tecnológico. A tecnologia da computação é um componente explícito nas definições fundadoras da Ciência da Informação, que investiga o fluxo da informação. Conforme Ferneda (2003), os sistemas de RI são a materialização dessa investigação, aplicando métodos computacionais para gerenciar o fluxo informacional. Esse processo, conforme ilustrado por Ferneda (2003, p. 15), utiliza uma busca que compara determinadas informações armazenadas com a expressão de busca formulada pelo usuário.

Aprofundando a arquitetura desse fluxo, é necessário primeiro compreender as finalidades basilares desses sistemas. Segundo Souza (2006), as principais funções de um Sistema de Recuperação da Informação (SRI) são: o armazenamento de documentos, a descrição das informações neles contidas e a recuperação das informações e dos próprios

documentos. Para operacionalizar essas funções, Cesarino (1985) define os SRI como um conjunto de operações consecutivas que visam aumentar a transmissão de informação relevante. O ponto central desta abordagem de Cesarino é que existem duas etapas distintas: primeiro, uma análise conceitual e, em seguida, uma tradução. A análise conceitual é o estágio intelectual de interpretação do conteúdo, enquanto a "tradução" é o ato de mapear esses conceitos para o vocabulário do sistema.

Aplicando esta arquitetura ao objeto deste estudo, a Base de efeitos TRIZ Multilíngue (Zaniro e Quoniam, 2024) em seu formato original (planilha) pode ser entendida como o resultado de uma "análise conceitual" (Cesarino, 1985), na qual seus autores já interpretaram e estruturaram o conteúdo. No entanto, ela carece da segunda etapa: uma tradução eficaz para um vocabulário de sistema que permita a comparação dinâmica com a expressão de busca do usuário (Ferneda, 2003). Nesta perspectiva, o desenvolvimento do artefato de mediação proposto nesta pesquisa constitui o processo de tradução. São estas operações consecutivas (Cesarino, 1985) que transformam a estrutura conceitual da planilha em um sistema de recuperação da informação, capaz de gerenciar o fluxo e entregar informação potencial (Ferneda, 2003) ao usuário.

O ponto de encontro onde o usuário expressa sua busca e o sistema entrega os resultados é o domínio da Interação Humano-Computador (IHC). Esta disciplina foca o projeto, avaliação e implementação de sistemas computacionais interativos para uso humano, bem como os fenômenos que cercam esse uso (Nielsen, 1994). No contexto da RI, a IHC é a disciplina responsável por projetar a interface que viabiliza o fluxo informacional. Conforme Nielsen (1994), um pilar central desta área é a engenharia de Usabilidade, que se concentra em garantir que esta interface seja eficiente, fácil de usar e subjetivamente satisfatória, permitindo que o usuário de fato concretize a recuperação da informação relevante.

Este conceito de eficiência é aprofundado por Steve Krug, cuja primeira lei da Usabilidade pode ser resumida na máxima: "Não me faça pensar". Esta abordagem defende que a interface deve ser autoevidente, intuitiva e exigir o mínimo de esforço cognitivo do usuário. Krug (2006) argumenta que os usuários não leem as páginas, eles as escaneiam, e tendem a satisfazer (escolher a primeira opção razoável que encontram) em vez de procurar a melhor opção. Para o design de um SRI, isso implica que a interface deve ser minimalista, eliminando distrações e palavras desnecessárias, permitindo que o usuário realize sua análise conceitual e execute a tradução (busca) com o menor número de etapas possível e sem ambiguidades.

O gerenciamento informacional contemporâneo tem sido ampliado pela Inteligência Artificial. Russell e Norvig (1995) definem a IA como o campo voltado ao estudo e à construção

de agentes racionais, concebidos para perceber as nuances de seu ambiente e executar ações que maximizem as chances de sucesso em um determinado objetivo. No escopo da RI, essa racionalidade traduz-se na capacidade de compreender consultas complexas e identificar relevâncias que escapam às buscas exatas por palavras-chave. Esse avanço foi viabilizado pela transição para modelos baseados em grandes volumes de dados que, conforme Goodfellow et al. (2016), permite que os sistemas computacionais aprendam representações de dados com múltiplos níveis de abstração, superando as limitações de vocabulários estritos criados manualmente.

Essa capacidade de abstração fornecida pela IA alcançou um novo patamar tecnológico com a introdução da arquitetura *Transformer*, base dos Grandes Modelos de Linguagem (LLM - *Large Language Models*) atuais, que permite ponderar a relevância e o contexto de uma consulta de forma global e simultânea. Contudo, a aplicação isolada desses modelos generativos pode resultar em respostas imprecisas, fenômeno conhecido como alucinação. Para mitigar esse problema e unir o potencial da IA à exatidão da RI, consolidou-se o paradigma da Geração Aumentada por Recuperação (RAG - *Retrieval-Augmented Generation*). Conforme proposto por Lewis et al. (2020), o RAG atua vinculando o modelo generativo a uma base de conhecimento externa e confiável. Essa abordagem atua como um motor avançado para refinar a análise conceitual descrita por Cesarino (1985), garantindo que a tradução da necessidade do usuário seja semanticamente fluida e ancorada no conhecimento estruturado do repositório.

Portanto, este capítulo construiu o arcabouço teórico que fundamenta a pesquisa. Partiu-se da definição da Ciência da Informação como a disciplina que estuda a transferência e o fluxo informacional (Barreto, 1996). Definiu-se o objeto (Informação Tecnológica/TRIZ) e a filosofia de acesso (Ciência Aberta), demonstrando a necessidade de guias práticos (Princípios FAIR) para mover os dados da mera disponibilidade para a acessibilidade efetiva. Por fim, exploraram-se as disciplinas técnicas para essa solução: a Recuperação da Informação e a Inteligência Artificial como os motores da "tradução" (Cesarino, 1985; Vaswani et al., 2017), e a Interação Humano-Computador (Nielsen, 1994; Krug, 2006) como o veículo para a entrega intuitiva ao usuário.

3 MATERIAIS E MÉTODOS

A presente pesquisa caracteriza-se como empírica-descritiva-exploratória e aplicada. Segundo Prodanov (2012), a pesquisa aplicada é aquela que visa à produção de conhecimentos e artefatos que podem ser utilizados de forma prática para a solução de problemas específicos. Simultaneamente, classifica-se como empírica pois, conforme Gil (2002), baseia-se na experiência e na observação direta da realidade para a coleta de dados e comprovação de hipóteses, não se restringindo a formulações teóricas. O problema central, neste caso, é a barreira de transferência da informação (Barreto, 1996) identificada na Base de efeitos TRIZ Multilíngue (Zaniro; Quoniam, 2024), e o artefato desenvolvido é a solução proposta.

Para a concepção deste artefato, o percurso metodológico foi estruturado em fases, iniciando-se por uma abordagem descritiva e exploratória. Na pesquisa descritiva, segundo Prodanov (2012), não há interferência nos dados, apenas são observados, analisados e registrados. A perspectiva exploratória é devido à expansão do assunto abordado, complementando com nova abordagem (Prodanov, 2012).

No aspecto exploratório, foi conduzida uma revisão bibliográfica utilizando as bases de dados Scopus e *Web of Science*, a partir de termos-chave como “princípios FAIR”, “Ciência Aberta”, “Repositório TRIZ”, “Base de Dados de Efeitos TRIZ”, “Informação Tecnológica”, “Sistemas Web” e “Inteligência Artificial”, sem restrição de períodos. Foram selecionados apenas artigos científicos, dissertações e teses.

No escopo da etapa exploratória, a inteligência artificial foi empregada como instrumento de apoio metodológico à pesquisa. Utilizou-se o *chatbot* Gemini (GOOGLE, 2023) para auxiliar no mapeamento teórico, facilitando a identificação de autores clássicos e contemporâneos correspondentes a conceitos e temas específicos abordados no estudo. Adicionalmente, a mesma ferramenta foi adotada na fase de sistematização do documento para auxiliar na padronização e formatação das referências bibliográficas, garantindo o rigor e a conformidade com as diretrizes da Associação Brasileira de Normas Técnicas (ABNT).

Seguindo o delineamento descritivo, a etapa inicial consistiu no acesso e análise do recurso informacional intitulado "Base de dados de efeitos TRIZ multilíngue ligada com IPC e CNAE", hospedado no Repositório Institucional da Universidade Federal de São Carlos (UFSCar). A análise dos metadados permitiu caracterizar a fonte primária, publicada em 25 de abril de 2024 por Zaniro e Quoniam, como um recurso alinhado aos princípios da ciência aberta, resultante da elaboração de algoritmos e consulta a APIs.

A descrição do recurso e suas palavras-chave associadas, TRIZ, Wikipédia, IPC, Patentes, CNAE, Inovação frugal e Algoritmos, confirmam a natureza interdisciplinar dos dados, que buscam conectar processos inventivos (TRIZ), propriedade industrial (Patentes/IPC) e setores econômicos (CNAE) em múltiplos idiomas.

Na verificação técnica, realizou-se a obtenção do arquivo primário, identificado como *triz_multilingue_total.xlsx* (com volume de aproximadamente 5.39 MB), acessível através do identificador persistente (URI) <https://repositorio.ufscar.br/handle/20.500.14289/20138>. Um aspecto indispensável identificado nesta etapa foi a licença de uso atribuída ao objeto: *Creative Commons Attribution-ShareAlike 3.0 Brazil* (CC BY-SA 3.0). Esta licença é fundamental para a viabilidade legal desta pesquisa, pois autoriza explicitamente o compartilhamento e a adaptação (criação de obras derivadas) do material, desde que citada a fonte e mantida a licença, legitimando o processo de reestruturação e modelagem proposto.

Com o arquivo recuperado, procedeu-se à inspeção interna dos dados. A análise quantitativa focou na mensuração do volume informacional, contabilizado pelo número de entradas de efeitos registradas. Simultaneamente, a análise qualitativa dedicou-se à identificação e conceituação dos atributos (colunas) que compõem a estrutura da base. Ainda no aspecto qualitativo, foi possível identificar e compreender os elementos multilíngues utilizados na descrição dos efeitos, bem como mapear as conexões e relações existentes entre eles. A fim de sistematizar a compreensão desta estrutura e fundamentar a etapa posterior de modelagem, os atributos identificados e suas respectivas definições foram condensados em uma tabela com os descritivos das colunas e os mapeamentos de idiomas.

Uma vez compreendida a estrutura física e semântica da base, avançou-se para a etapa de diagnóstico de conformidade. Nesta fase, o recurso informacional foi submetido a uma avaliação baseada nos Princípios FAIR, conforme fundamentado teoricamente por Wilkinson *et al.* (2016) e detalhado na Seção 2 deste trabalho.

A avaliação consistiu na verificação individual de cada um dos princípios FAIR frente às características observadas tanto no Repositório Institucional (metadados e identificadores) quanto no arquivo de dados (estrutura interna da planilha). O objetivo deste procedimento não foi apenas classificar a base, mas identificar lacunas de interoperabilidade e acessibilidade que justificassem intervenções técnicas.

Esta análise diagnóstica serviu como base para a definição dos requisitos do novo artefato, permitindo mapear quais melhorias e transformações seriam necessários na etapa de modelagem e desenvolvimento para que a base passasse a atender mais critérios de uma gestão de dados aberta.

Complementando o diagnóstico interno, conduziu-se uma pesquisa exploratória em bases correlatas disponíveis na *Web*, com o objetivo de identificar padrões de interface, estratégias de recuperação da informação e lacunas funcionais que pudessem balizar o desenvolvimento do novo artefato. Foram selecionadas para análise duas plataformas de referência na área: a *Aulive* e a *Oxford Effects Database*. A escolha dessas fontes específicas justifica-se pela sua relação com o objeto de estudo, visto que ambas serviram como fontes de efeitos TRIZ para a estruturação da Base de efeitos TRIZ Multilíngue proposta por Zaniro e Quoniam (2024).

Na análise da plataforma *Aulive* (2025), observou-se um modelo de recuperação estruturado em torno de dois eixos: a "função" (ação) e o "estado" (objeto). Verificou-se que este sistema entrega resultados acompanhados de recursos didáticos, como animações e exemplos de aplicação, o que sugere a importância de elementos visuais na mediação da informação tecnológica. Contudo, identificou-se como limitação a restrição do conteúdo exclusivamente ao idioma inglês e a dependência exclusiva de parâmetros pré-definidos, que restringe o usuário a uma única estratégia de consulta.

Em seguida, examinou-se a *Oxford Effects Database* (2025), em que se constatou uma maior flexibilidade na formulação das consultas. A plataforma oferece três paradigmas de busca — por "função", "parâmetro" ou "transformação" —, permitindo ao usuário diferentes estratégias para expressar seu problema inventivo. Um ponto importante observado foi a estratégia de indexação: os resultados funcionam como um índice remissivo, direcionando o usuário para artigos da Wikipédia (em inglês) para o aprofundamento do conteúdo. Esta característica valida a abordagem de uso de fontes colaborativas, similar à base objeto deste estudo, porém, assim como a anterior, limita-se ao idioma inglês.

Esta etapa foi determinante para a concepção do sistema proposto, pois evidenciou a necessidade de manter a flexibilidade de busca (modelo da *Oxford*) e a clareza na apresentação (modelo da *Aulive*), mas reforçou a lacuna crítica de acessibilidade linguística que o presente trabalho visa solucionar através da abordagem multilíngue.

Fundamentado no diagnóstico inicial e na análise das plataformas correlatas, avançou-se para a fase de concepção e desenvolvimento do artefato de mediação. A premissa metodológica para esta etapa foi a aplicação prática dos princípios FAIR como diretriz de arquitetura. O objetivo central da modelagem foi estruturar e relacionar os diversos elementos que compõem a base de efeitos TRIZ multilíngue, transformando a estrutura plana original em uma representação sistêmica capaz de preservar a totalidade das relações semânticas de cada efeito.

Para operacionalizar a interoperabilidade (princípio *Interoperable*), optou-se pela modelagem dos dados em uma estrutura que estabelecesse o relacionamento entre todos os componentes da base de efeitos TRIZ. A construção deste modelo derivou da análise minuciosa da estrutura original e das percepções de Usabilidade obtidas no estudo das plataformas *Aulive* e *Oxford*.

O núcleo central da modelagem reside na entidade *Triz Effect*, que atua como o elemento conector de todas as tabelas descritivas. Cada registro nesta tabela representa um efeito único, associado às tabelas descritivas, além de armazenar atributos de qualificação como o *IPC_Score* (que quantifica a relação com a classificação de patentes) e indicadores booleanos que sinalizam a presença do efeito nas bases externas analisadas.

As características do efeito foram decompostas em tabelas independentes, com o objetivo de descreverem o efeito cada qual a seu modo, para garantir a integridade e flexibilidade das consultas: *Type* (tipos), *Task* (ações), *Mode* (modos de aplicação) e *Target* (alvos). O conteúdo descritivo textual, incluindo definições e *links* multimídia, foi normalizado na tabela *Term*. Destaca-se ainda a integração com tabelas de vocabulários controlados externos, CNAE e IPC, que vinculam os efeitos a contextos industriais e tecnológicos, enriquecendo a representação da informação.

Um desafio metodológico foi o requisito de representação multilíngue. A decisão técnica para endereçar esta questão sem redundância de dados foi a adoção do tipo de dado JSONB (*Binary JavaScript Object Notation*) do sistema gerenciador de banco de dados PostgreSQL, tecnologia baseada em código aberto. Esta abordagem permitiu que termos mapeados em quatro idiomas (português, inglês, francês e espanhol) fossem armazenados em um único registro estruturado, facilitando a internacionalização. Por exemplo, para a ação "Mover", o atributo *task_name* armazena a estrutura: {"pt": "Mover", "en": "Move", "fr": "Déplacer", "es": "Mover"}. O uso de JSONB permite indexação e consultas eficientes dentro do próprio documento, permitindo que a recuperação da informação opere de forma transparente em relação ao idioma selecionado pelo usuário.

Paralelamente à modelagem, realizou-se o processamento e normalização das ilustrações dos efeitos. Originalmente, as ilustrações encontravam-se em formato HTML (*Hypertext Markup Language*), o que dificultava sua pré-visualização imediata em listas de resultados. Para solucionar esta questão de acessibilidade, desenvolveu-se um *script* de automação em linguagem *Python*, que converteu massivamente os arquivos originais para o formato GIF (*Graphics Interchange Format*). Esta conversão permitiu que o conteúdo visual

pudesse ser exibido nativamente em qualquer navegador, ampliando a compatibilidade do sistema.

O artefato passou por uma etapa de enriquecimento de dados com a incorporação das *Catchwords* (índice de palavras-chave da IPC). É importante destacar que esse vocabulário não integrava a base de efeitos TRIZ original. Para viabilizar sua utilização, os dados foram obtidos diretamente do portal da WIPO em formato XML. O processamento desse material foi realizado por meio de um *script* desenvolvido em Python, responsável por ler o arquivo, extrair as palavras-chave e inseri-las em tabelas estruturadas no banco de dados PostgreSQL. Como o arquivo XML original já continha os respectivos códigos IPC atrelados a cada *Catchword*, foi possível estabelecer um relacionamento direto com os números IPC preexistentes na base de dados relacional.

Com os dados estruturados e os ativos processados, procedeu-se ao desenvolvimento do artefato de mediação, projetada para traduzir a complexidade do modelo relacional em uma experiência de recuperação da informação intuitiva. Utilizou-se o *framework Laravel* (PHP: *Hypertext Preprocessor*) para a construção do processamento. A escolha justifica-se pela robustez na criação de recursos. Para a camada de Interação Humano-Computador, adotou-se o *framework Vue.js*. Esta tecnologia foi selecionada por sua arquitetura baseada em componentes reativos, permitindo a construção de uma interface de busca dinâmica que não exige o recarregamento da página a cada filtro aplicado, otimizando a experiência do usuário.

A lógica operacional do artefato de mediação é regida por algoritmos, que constituem o método essencial para transformar a interação do usuário em resultados tangíveis. Segundo Cormen (2002), um algoritmo define-se como qualquer procedimento computacional bem definido que toma um conjunto de valores como entrada e produz um conjunto de valores como saída, operando por meio de uma sequência de passos computacionais que transformam o *input* no *output*. No contexto deste artefato, a construção algorítmica atua como o mecanismo central de mediação: é ela a responsável por processar os parâmetros de busca complexos inseridos pelo usuário (entrada) e percorrer a estrutura de dados relacional e JSONB para recuperar, com precisão e eficiência, os efeitos TRIZ correspondentes (saída). Portanto, os algoritmos não representam apenas um recurso técnico de programação, mas a materialização da estratégia de recuperação da informação que viabiliza o acesso ordenado ao conhecimento da base de efeitos TRIZ.

A infraestrutura de hospedagem para a Base de Dados e o ambiente do artefato foi estabelecida em um ambiente *cPanel* padronizado. A escolha por este ambiente reforçou a decisão metodológica de utilizar o *framework Laravel* para a construção da API devido à sua

compatibilidade nativa com ambientes baseados em PHP, como é o caso do *cPanel*. Desta forma, as escolhas de software e infraestrutura foram mutuamente alinhadas para garantir a estabilidade e a integração da base de dados PostgreSQL com a interface dinâmica em Vue.js.

Além das abordagens tradicionais de busca, projetou-se a implementação de uma camada de recuperação semântica fundamentada na técnica de RAG. Para viabilizar essa funcionalidade, foi necessária a adoção de um banco de dados vetorial, capaz de processar buscas por similaridade espacial. Inicialmente, avaliou-se o uso da extensão *pgvector* para o PostgreSQL, bem como soluções como *ChromaDB* e *Milvus*. Contudo, as restrições de instalação no núcleo do servidor *cPanel*, a exigência de infraestrutura dedicada e a alta complexidade computacional inviabilizaram essas opções. A solução metodológica adotada foi o *Pinecone*, um banco de dados vetorial que permitiu a integração via API sem sobrecarregar o ambiente de hospedagem compartilhado.

Para a orquestração da inteligência artificial, conduziu-se uma análise comparativa de LLMs. Modelos como o GPT-4 (OpenAI) apresentaram custos elevados para o volume de dados e limitações de taxa de requisição, enquanto o Llama 3 (Meta), apesar de possuir código aberto, demandaria uma infraestrutura de hardware baseada em GPUs (do inglês *Graphics Processing Units*, unidades de processamento gráfico projetadas para executar cálculos matemáticos massivos em paralelo) de alto desempenho, incompatível com o servidor utilizado. Diante disso, optou-se pela adoção dos modelos da família Gemini (GOOGLE, 2023), que se destacaram pela superioridade no processamento de textos multilíngues e pela integração nativa com modelos de *embedding*, uma técnica que converte o conteúdo textual em representações numéricas (vetores), permitindo que o sistema interprete e relacione matematicamente o sentido semântico das palavras.

A operacionalização do RAG foi dividida em etapas estruturadas no processamento com Laravel. Na fase de indexação, os dados descritivos em português, inglês, espanhol e francês foram concatenados e processados em lotes pelo modelo *gemini-embedding-001*, gerando vetores de 768 dimensões que foram armazenados no *Pinecone*. No fluxo de consulta, o sistema emprega o modelo Gemini 2.5 Flash Lite, configurado com baixa temperatura (0.1), parâmetro técnico que reduz a aleatoriedade da geração de texto, para extrair as entidades ("Tarefa" e "Alvo") da pergunta em linguagem natural, reduzindo ruídos antes da busca vetorial.

Em seguida, os registros técnicos mais similares recuperados pelo banco vetorial são enviados ao modelo Gemini 1.5 Flash (com temperatura de 0.4), responsável por sintetizar a resposta final. Esse processo opera sob diretrizes estritas de *prompt*, restringindo a IA a utilizar exclusivamente o contexto técnico recuperado, de modo a mitigar alucinações e garantir que a

resposta mantenha a exatidão e a terminologia multilíngue da base original. Desta forma, as escolhas de software, infraestrutura e inteligência artificial foram mutuamente alinhadas para garantir a estabilidade e a integração da base de dados PostgreSQL e do banco vetorial com a interface dinâmica em Vue.js.

A arquitetura do sistema foi estruturada para oferecer quatro estratégias principais de recuperação, definindo os requisitos funcionais da aplicação:

1. Recuperação por Problema (Perspectiva Funcional): Inspirada nos paradigmas clássicos da TRIZ, a funcionalidade foi programada para permitir ao usuário expressar sua necessidade através de três filtros distintos: por Função (Ação + Objeto), por Parâmetro (Operação + Parâmetro) ou por Transformação de estados.
2. Recuperação por Sugestão TRIZ: Implementou-se um mecanismo de busca direta nominal, permitindo a localização do efeito pelo seu nome científico ou inventivo.
3. Recuperação por Perspectiva Industrial (CNAE): O sistema integrou a busca pela Classificação Nacional de Atividades Econômicas. Esta funcionalidade utiliza a tabela de relacionamento modelada para recuperar efeitos aplicáveis a setores econômicos específicos.
4. Recuperação por Perspectiva Patentária (IPC): Desenvolveu-se o recurso de busca pela Classificação Internacional de Patentes, permitindo conectar soluções inventivas presentes em documentos de patentes aos efeitos científicos capazes de solucionar problemas técnicos.

Por fim, como etapa de validação metodológica do processo de desenvolvimento, o artefato foi submetido a uma revisão de conformidade com os princípios FAIR, visando assegurar que as funcionalidades implementadas respondessem às barreiras de transferência identificadas.

A apresentação do artefato, bem como a discussão sobre sua usabilidade e impacto na transferência da informação, serão detalhadas na seção de Resultados.

4 RESULTADOS

Esta seção materializa o artefato proposto bem como as atividades realizadas com base nos processos metodológicos descritos.

4.1 ANÁLISE DIAGNÓSTICA DA BASE DE EFEITOS TRIZ MULTILÍNGUE

A base de efeitos TRIZ foi analisada para compreender suas características e contexto. Foram identificadas 47 colunas, descritas na Tabela 1.

Tabela 1: Colunas da Base de efeitos TRIZ

Título da coluna da base de efeitos TRIZ	Descrição	Equivalen te em português	Equivalen te em espanhol	Equivalen te em francês
Oxford	Indica se o registro está presente na base de conhecimento Oxford.			
Aulive	Indica se o registro está presente na base de conhecimento Aulive.			
Wikilink	Indica se há um <i>link</i> associado ao termo na Wikipédia.			
Mode	Representa o “modo” como um conceito é aplicado.	Mode PT	Mode ES	Mode FR
Type	Indica a natureza do conceito.	Type PT	Type ES	Type FR
Task	Detalha a tarefa ou ação prática relacionada ao conceito descrito, especificando a sua utilidade ou função.	Taskt PT	Task ES	Task FR
Target	Especifica o alvo ou objeto pretendido para a aplicação do	Target PT	Target ES	Target FR

	conceito, tornando claro o foco da funcionalidade			
Name	Nome do efeito, relacionado ao par tarefa-alvo (task-target)	PT Term	ES Term	FR Term
Description	Descrição detalhada do conceito com contexto e exemplos.	PT Description	ES Description	FR Description
Note	Observações e detalhes extras sobre o conceito.			
Example	Exemplo de aplicação do conceito.			
EN <i>Link</i>	<i>Link</i> que redireciona à Wikipédia sobre o conceito.	PT <i>Link</i>	ES <i>Link</i>	FR <i>Link</i>
Nota TRIZ-IPC (0-5)	Valor quantitativo, entre 0 e 5, que representa o grau de confiabilidade da ligação entre efeitos TRIZ e códigos IPC			
IPC	Código IPC.			
ID Seção CNAE	ID da Seção CNAE			
Seção CNAE	Descrição Nominal da Seção CNAE			
ID Divisão CNAE	ID da Divisão CNAE			
Divisão CNAE	Descrição Nominal da Seção CNAE			
ID Grupo CNAE	ID do Grupo CNAE			
Grupo CNAE	Descrição Nominal da Seção CNAE			

ID Classe CNAE	ID da Seção CNAE
Classe CNAE	Descrição Nominal da Seção CNAE
ID Subclasse CNAE	ID da Subclasse CNAE
Subclasse CNAE	Descrição Nominal da Subclasse CNAE
GIF Animation	Imagem animada para representar o conceito.
Animation <i>Link</i>	<i>Link</i> que redireciona para a imagem animada.

Fonte: Elaborado pelo autor

A análise descritiva e qualitativa da base, sistematizada na Tabela 1, confirmou a complexidade estrutural do objeto de estudo. Observou-se que a base é composta por quatro grupos de atributos centrais. O primeiro grupo é o Descritivo-Semântico, em que colunas como *Name*, *Description*, *Type*, *Task* e *Target* representam a estrutura de conhecimento da metodologia TRIZ e seus equivalentes em quatro idiomas (português, inglês, francês e espanhol).

O segundo grupo é o de Ligação Externa, em que as colunas IPC, Nota TRIZ-IPC (0-5) e os oito campos referentes ao código CNAE estabelecem o mapeamento entre o efeito científico e os contextos industrial e patentário. O terceiro grupo é o de Mapeamento Colaborativo, composto pelos *links* da Wikipédia e indicadores de presença (*Oxford* e *Aulive*). Finalmente, o quarto grupo é o de Conteúdo Multimídia e Metadados, que inclui os campos para GIF Animation e Note. Essa riqueza de atributos e a complexidade hierárquica das colunas de classificação (CNAE) justificaram o desenvolvimento do modelo de dados proposto, necessário para otimizar a recuperação dessas conexões.

A inter-relação funcional entre esses grupos pode ser exemplificada pelo registro do efeito 'Absorção'. No nível Descritivo-Semântico, o conceito é definido pela função TRIZ 'Extrair' aplicada ao alvo 'Gás'. Essa definição conecta-se à classificação de patentes (Ligação Externa) através do código IPC B01D53/14 — específico para processos de separação de gases —, que possui nota máxima de confiabilidade na base. Simultaneamente, o código da Subclasse

CNAE 20.14-2/00 ('Fabricação de gases industriais') associa o efeito científico ao setor industrial correspondente. Por fim, o Mapeamento Colaborativo amplia o contexto via *links* multilíngues (como pt.wikipedia.org/wiki/Absorção), demonstrando como a estrutura da base articula o conceito abstrato com sua aplicação prática e sua categorização formal.

4.2 AVALIAÇÃO DIAGNÓSTICA FAIR

Esta subseção apresenta os achados da Avaliação Diagnóstica realizada sobre a Base de efeitos TRIZ em seu formato original (planilha), utilizando os Princípios FAIR como protocolo de conformidade, conforme estabelecido na metodologia (Seção 3). O objetivo central desta etapa foi validar, com base em diretrizes internacionais de Ciência Aberta (Wilkinson *et al.*, 2016), as lacunas de acessibilidade e interoperabilidade do recurso, confirmando a existência de uma barreira de transferência que impedia o uso eficaz dos dados. A análise sistemática de cada princípio forneceu a base empírica para a definição dos requisitos funcionais e de modelagem, delineando as transformações necessárias para que o novo artefato pudesse concretizar o fluxo informacional.

O princípio *Findable* é identificado com base nas colunas listadas na Tabela 1. O identificador único previsto pelo princípio está relacionado às colunas de ID Divisão CNAE, ID Grupo CNAE, ID Classe CNAE, ID Subclasse CNAE que, juntas, relacionam o efeito TRIZ a um único CNAE. Além disso, a descrição rica de metadados também é encontrada por meio das colunas que descrevem o efeito TRIZ, isto é, *Name*, *Description*, *Mode* e *Type*. Os *links* disponíveis em português, inglês, espanhol e francês permitem conectar os dados e seus identificadores de forma clara, considerando respectivamente as colunas *PT Link*, *EN Link*, *ES Link*, *FR Link*.

Ainda sobre o princípio *Findable*, foram encontradas colunas que relacionam diretamente o efeito TRIZ à sua respectiva página na Wikipédia. Estes *links* permitem associação com a Wikipédia nos quatro diferentes idiomas, entretanto, é importante ressaltar que nem todos os efeitos TRIZ possuem estas colunas preenchidas com *links*, o que torna este requisito parcialmente atendido. Estes *links*, quando preenchidos, permitem maior acessibilidade e expansão dos contextos dos efeitos TRIZ listados, uma vez que a Wikipédia fornece um suporte enciclopédico e permite outras pesquisas.

No que se refere ao princípio *Accessible*, é possível dizer que a base de efeitos TRIZ atende partes do princípio. Atende porque possui campos descritivos que indicam o uso de padrões estabelecidos na comunidade científica para descrever informações inerentes ao efeito TRIZ representado em cada linha, conforme as colunas *Name*, *Description*, *Mode* e *Type*

descrevem nos quatro idiomas identificados. Em complemento, os *links* presentes na base, quando preenchidos, utilizam protocolos amplamente utilizados, como HTTP (*Hypertext Transfer Protocol*), o que promove a universalidade de acesso em diferentes sistemas.

Neste princípio, também é possível mencionar que alguns conceitos associados aos efeitos TRIZ indicam que, caso as colunas principais que descrevem o efeito TRIZ não estejam disponíveis, os metadados relacionados a ele continuam disponíveis como referência acessível ou futura recuperação. Os *links* que direcionam para a Wikipédia garantem que dados complementares ao efeito TRIZ possam ser acessados de forma independente ao efeito, porém, nem todos os efeitos TRIZ possuem este *link* preenchido.

Portanto, no que se refere à acessibilidade (*Accessible*), a base de efeitos TRIZ reflete parcialmente este princípio, porque há lacunas e áreas que podem ser otimizadas. Embora alguns protocolos universais estejam presentes em URLs, a base está disponível em planilha eletrônica, o que não permite identificar com clareza se implementa protocolos universais que permitem que a planilha seja acessada amplamente por sistemas diversos. Ainda neste sentido, não é possível identificar se a base utiliza mecanismos de autenticação para segurança de acessos, mas entende-se que estar disponibilizada em um repositório institucional supre a necessidade de autenticação. Apesar de possuir dados e metadados com vocabulários comuns à comunidade científica, não está especificado claramente qual padrão é utilizado.

A base de efeitos TRIZ atende o princípio da interoperabilidade (*Interoperable*) no quesito de linguagem formal e acessível, como as colunas que constituem o CNAE. O CNAE pode ser considerado um vocabulário padronizado para contextos industriais no Brasil e, ao estar referenciado na base, permite que os efeitos TRIZ sejam associados a outros contextos que utilizem o CNAE como identificador. Por outro lado, apesar de existirem colunas estruturadas e bem definidas, não é possível identificar uma representação formalizada como formato RDF (*Resource Description Framework*) ou similares. Assim, não está totalmente apta a se conectar com outras fontes de dados por não seguir padrões amplamente aplicáveis.

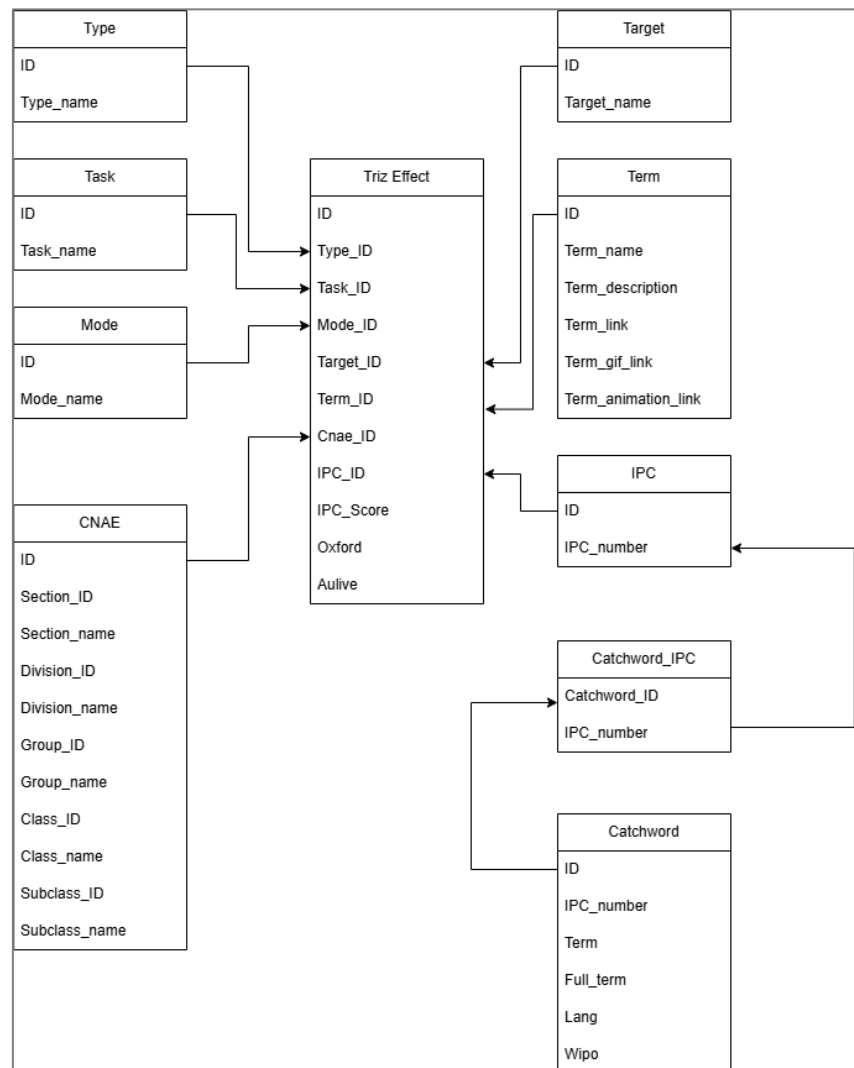
Sobre o princípio de dado reutilizável (*Reusable*), é possível notar que as colunas presentes na base descrevem detalhadamente os efeitos TRIZ e suas características, além de disponibilizar o conteúdo em português, inglês, espanhol e francês. No repositório institucional em que a base de efeitos TRIZ é disponibilizada, é explicitado que a base possui a licença *Creative Commons Attribution-ShareAlike 3.0 Brazil* (CC BY-SA 3.0), que prevê a utilização dos dados de forma pública, desde que haja pleno crédito aos autores. Os dados, conforme citado anteriormente, seguem alguns padrões comuns na comunidade científica, e outros não são identificados, o que torna este princípio parcialmente atendido.

A análise empírica revelou que a base de dados, estruturada em 47 colunas com descrições multilíngues, metadados organizacionais e códigos CNAE, apresenta uma adesão apenas parcial aos princípios FAIR. Embora o princípio *Findable* seja favorecido pela presença de identificadores únicos e descrições ricas, a incompletude dos *links* para a Wikipédia compromete a plena conectividade e a localização dos dados. De modo análogo, os princípios *Accessible*, *Interoperable* e *Reusable* enfrentam limitações estruturais: apesar da utilização de protocolos padrão (HTTP) e da adoção de uma licença *Creative Commons* — que viabiliza o uso público —, a ausência de padrões formais de interoperabilidade e a restrição ao formato de planilha eletrônica indicam a necessidade de aprimoramentos técnicos para garantir a total conformidade com as normas da comunidade científica e a efetiva integração sistêmica.

4.3 ESTRUTURAÇÃO DAS TABELAS DO ARTEFATO DE MEDIAÇÃO

A subseção apresenta a implementação da estrutura das tabelas, que constitui a tradução (Cesarino, 1985) da estrutura conceitual da planilha original para um formato sistêmico. O desenvolvimento deste esquema é o resultado direto da avaliação diagnóstica (Seção 4.2), visando ampliar a aderência aos princípios FAIR e viabilizar a integridade e flexibilidade necessárias para a recuperação da informação. O modelo implementado, ilustrado na Figura 1, adota uma arquitetura centralizada.

Figura 1: Visão Geral das Tabelas do Artefato de Mediação



Fonte: Elaborado pelo autor

O núcleo central da modelagem reside na entidade *Triz Effect*. Ela atua como a tabela fato do sistema, centralizando todas as ligações descritivas e classificatórias através de chaves estrangeiras. A *Triz Effect* armazena metadados de qualificação essenciais, como *IPC_Score* (que quantifica a confiabilidade da ligação com patentes) e os indicadores booleanos *Oxford* e *Aulive*, que sinalizam a presença do efeito nas bases correlatas.

As entidades *Task*, *Type*, *Mode* e *Target* funcionam como tabelas descritivas que desmembram as características do efeito TRIZ, viabilizando a integridade dos dados e otimizando as consultas multifacetadas da aplicação. A entidade *Term* é vital, pois armazena os metadados descritivos e *links* externos, enquanto as entidades *IPC* e *CNAE* representam os vocabulários controlados externos que enriquecem o modelo. A *CNAE* é modelada com sua estrutura hierárquica completa (Seção, Divisão, Grupo, Classe, Subclasse) para permitir

consultas precisas por contexto industrial, tal como requisitado pelas estratégias de busca (Seção 3).

É fundamental ressaltar que a característica multilíngue da base foi metodologicamente endereçada no nível de implementação do banco de dados. Para as colunas de conteúdo descritivo (como os campos *Task_name*, *Target_name* e os atributos da tabela *Term*), adotou-se o tipo de dado JSONB do PostgreSQL, garantindo que a informação em português, inglês, francês e espanhol seja armazenada de forma estruturada e em um único registro, promovendo a interoperabilidade linguística e o uso eficiente do sistema de recuperação. É importante ressaltar que o tipo de dado JSONB é uma técnica de internacionalização de baixo nível (i18n) que permite a integridade dos dados, evita a criação de múltiplas colunas por idioma e permite consultas e indexação rápidas, tornando a recuperação da informação independente da língua selecionada.

Dessa forma, o modelo apresentado não apenas organiza os dados de maneira íntegra, mas também viabiliza as quatro estratégias de recuperação — funcional, de solução, industrial (CNAE) e patentária (IPC) — que serão materializadas na interface do artefato.

A etapa subsequente à modelagem lógica consistiu na implementação das tabelas no Sistema Gerenciador de Banco de Dados (SGBD) PostgreSQL. A descrição técnica detalhada das tabelas implementadas fornece a evidência da tradução dos conceitos da planilha para uma estrutura padronizada e legível por máquina, requisito fundamental do princípio FAIR de Interoperabilidade. Foi implementado um projeto conceitual em DDL (*Data Definition Language*).

O projeto DDL buscou resolver os desafios de integridade referencial e representação de dados complexos de três formas principais.

Para modelagem hierárquica e vínculos externos, A tabela *cnae* foi estruturada para armazenar a hierarquia completa da CNAE (Seção, Divisão, Grupo, Classe e Subclasse) em um formato normalizado. Esta abordagem garante que a recuperação da informação pelo contexto industrial seja precisa e que os dados de classificação não sejam redundantes. A tabela *ipc* armazena os códigos da IPC. A tabela central *triz_effects* conecta-se a estas classificações por chaves estrangeiras (*cnae_id*, *ipc_id*), o que confere ao sistema a capacidade de realizar buscas e filtros complexos por perspectivas industrial e patentária, conforme os requisitos funcionais da aplicação.

Aprofundando a modelagem dos vínculos externos, especificamente para o domínio patentário, o modelo incorporou as entidades *Catchword* e *Catchword_IPC*. A tabela *Catchword* foi projetada para armazenar o índice alfabético oficial da WIPO (disponibilizado

em inglês e francês), atuando como uma porta de entrada textual para as classificações de patentes. Visando otimizar a resposta na camada de interface e a compreensão do contexto pelo usuário, essa estrutura armazena não apenas o termo isolado, mas também o seu caminho hierárquico completo (*full_term*), preservando o aninhamento lógico originário dos dados da WIPO.

Para integrar este vocabulário textual à estrutura central de códigos, a tabela *Catchword_IPC* estabelece a ligação direta entre a palavra-chave e o símbolo correspondente da IPC (*IPC_number*). Essa associação permite que a recuperação da informação opere de forma flexível: o sistema cruza as buscas textuais do usuário pelas *catchwords* para localizar o código IPC, que, por sua vez, está associado à tabela *Triz Effect*, ampliando a encontrabilidade (*Findable*) das soluções tecnológicas na base.

Para a recuperação multifacetada, as tabelas *tasks* e *targets* foram projetadas para dar suporte direto às estratégias de busca funcional. A inclusão de campos booleanos (*flag_mode_function*, *flag_mode_parameter*, *flag_mode_transform*) nestas tabelas materializa as regras do sistema de recuperação, permitindo que o mecanismo de processamento e a interface identifiquem rapidamente quais tarefas e alvos são válidos para cada modo de consulta, otimizando a eficácia da recuperação da informação. A tabela *terms* armazena os dados descritivos (nome, descrição, *links*) e indicadores (*flag_en*, *flag_pt*, *flag_es*, *flag_fr*), permitindo que a interface apresente o conteúdo correto e utilize o *link* da Wikipédia no idioma selecionado pelo usuário, reforçando a acessibilidade linguística do recurso.

4.4 O ARTEFATO

Esta subseção apresenta o artefato de mediação desenvolvido, o resultado central da pesquisa aplicada, que viabiliza a consulta aos dados da base de efeitos TRIZ (Seção 4.3) em uma interface funcional.

O objetivo primordial deste artefato é superar a barreira de transferência da informação (Barreto, 1996), garantindo que o vasto estoque de dados da Base TRIZ se converta em um fluxo acessível e utilizável para o usuário final. Conforme estabelecido na metodologia, a concepção do artefato foi guiada pelos princípios da IHC e Usabilidade, visando traduzir a complexidade do modelo relacional em uma experiência intuitiva. A Figura 2 ilustra a interface de busca implementada.

Figura 2: Interface de Busca

Fonte: Elaborado pelo autor

Conforme ilustrado na Figura 2, a interface instancia os princípios de Usabilidade discutidos no referencial teórico, sendo projetada para reduzir o esforço cognitivo e a complexidade de navegação. A arquitetura foi concebida sob o princípio da simplicidade (Krug, 2006), apresentando todos os eixos de recuperação na mesma tela e agrupando as opções de filtro sob rótulos semânticos (Problema, Sugestão TRIZ, Perspectiva Industrial, Perspectiva Patentária). Esta organização minimiza a necessidade de trocas de página, otimizando o processo de busca do usuário. Estrategicamente, esta estrutura multifacetada é importante para a Recuperação da Informação, pois permite que o usuário expresse sua necessidade através de diferentes vocabulários — funcional (Função/Objeto), de solução (Efeito), econômico (CNAE) ou patentário (IPC) — maximizando a encontrabilidade dos dados e refletindo a complexidade do modelo relacional na interface de maneira intuitiva. O seletor de idioma reforça a acessibilidade linguística da base, controlando em tempo real o vocabulário utilizado na consulta.

Complementarmente, a interface explicita sua conexão com a Wikipédia logo em seu cabeçalho. O título da página, "Efeitos TRIZ", possui um *hiperlink* na sigla "TRIZ" que direciona o usuário para o artigo explicativo da metodologia na Wikipédia. O destino deste *link* é atualizado dinamicamente de acordo com a opção definida no seletor de idioma. A implementação deste recurso reitera uma característica conceitual do artefato: a abordagem multilíngue não se restringe a uma tradução literal de termos, mas atua como um mapeamento semântico e cultural.

A funcionalidade de Busca por Problema (Função) foi estabelecida como o modo de recuperação padrão do sistema, uma vez que representa o cerne da metodologia TRIZ, baseada na relação fundamental Tarefa-Alvo (ou Ação-Objeto). Conforme ilustrado na Figura 3, a interface foi projetada para otimizar a precisão da busca através da compatibilidade dinâmica

entre os campos Ação e Objeto. Ao selecionar uma Ação, o vocabulário de Objetos é imediatamente ajustado para exibir apenas as opções compatíveis, e vice-versa. Este mecanismo de restrição é uma estratégia importante de engenharia de Usabilidade, pois impede buscas sem resultados válidos, reduzindo a frustração do usuário e elevando a precisão da recuperação da informação. O campo Solução, por sua vez, é opcional e serve como um filtro avançado para refinar a consulta quando o usuário já possui um termo TRIZ conhecido. Os resultados são exibidos em formato de acordeão, permitindo ao usuário escanear e priorizar os Efeitos encontrados antes de expandir o registro para análise detalhada, promovendo a eficiência na transferência da informação.

Figura 3: Busca por Função

Efeitos TRIZ

Idioma: Português

Problema: ☒ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito

Perspectiva Industrial: ☐ CNAE ☐ IPC

Ação: Desmembrar Objeto: Sólido Solução: Buscar

124 Efeitos TRIZ Encontrados

- [Ablação](#)
- [Ablação por laser](#)
- [Abrasão](#)
- [Acoustic Vibration](#)
- [Adsorção](#)
- [Aerobic Digestion](#)

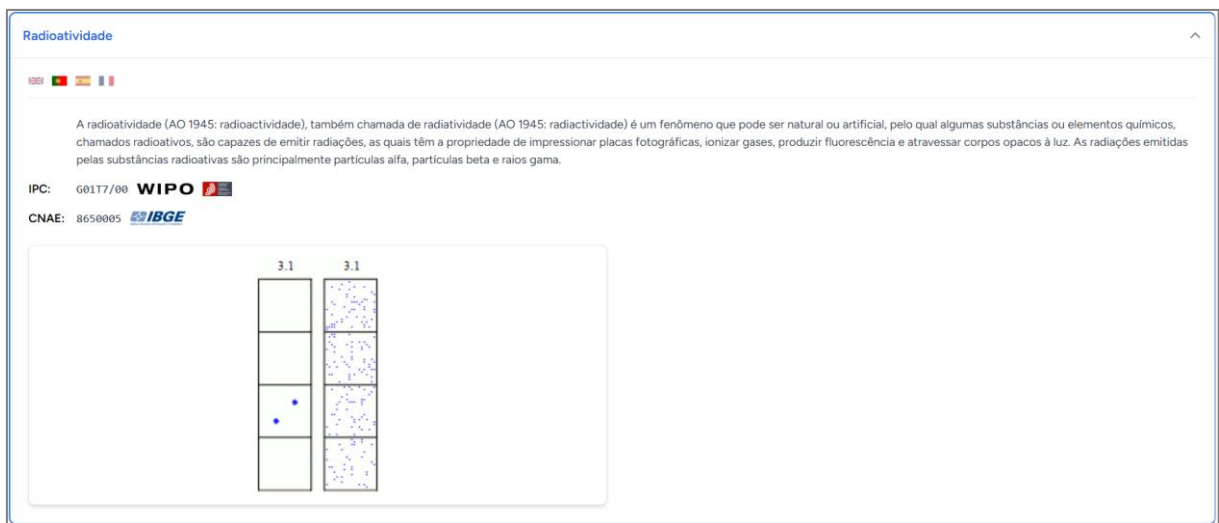
Fonte: Elaborado pelo autor

Ao expandir um dos itens da lista no formato acordeão, o sistema revela o detalhamento do efeito, projetado para maximizar a reusabilidade e a interoperabilidade da informação, conforme mostram as Figuras 4 e 5. O título do efeito, visível no cabeçalho do acordeão, é um *link* direto para a respectiva página da Wikipédia no idioma selecionado (português, inglês, espanhol ou francês), garantindo a acessibilidade ampliada à descrição do conceito. As bandeiras abaixo do título atuam como indicadores multilíngues, sinalizando rapidamente os idiomas para os quais há mapeamento disponível para aquele efeito específico.

O corpo do resultado apresenta a descrição detalhada do efeito, os recursos visuais disponíveis (como a imagem animada) e os vínculos de classificação especializados. Os códigos IPC e CNAE são apresentados como *links* externos para as bases de dados oficiais – WIPO,

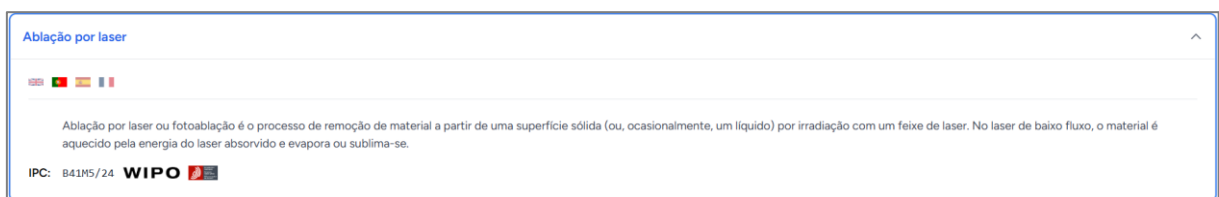
Espacenet (para patentes) e IBGE (para o CNAE) – transformando o artefato em um nó de informação que conecta o conhecimento inventivo a contextos econômicos e tecnológicos formais. Esta estratégia de ligação externa cumpre o princípio FAIR de Interoperabilidade, permitindo que o usuário navegue entre o conhecimento científico e as fontes de informação proprietária/oficial. É importante notar que o sistema exibe os dados (descrição, *links*) com o grau de detalhe e completude em que foram encontrados na base original, respeitando a integridade da fonte.

Figura 4: Resultado da Busca por Função (Radioatividade)



Fonte: Elaborado pelo autor

Figura 5: Resultado da Busca por Função (Ablação por laser)



Fonte: Elaborado pelo autor

Uma estratégia adicional foi implementada para maximizar a Encontrabilidade (*Findable*) e mitigar a frustração do usuário em casos de lacunas linguísticas. Embora a busca seja feita no idioma selecionado (Ex.: português), o sistema utiliza um mecanismo de contingência multilíngue. Conforme demonstrado na Figura 6, se o efeito TRIZ correspondente não possuir mapeamento para o idioma primário (português), o resultado será exibido no próximo idioma mapeado (geralmente o Inglês, como no caso de "*Acoustic Vibration*"). Esta estratégia de degradação prioriza a transferência do conhecimento (Barreto, 1996) em

detrimento da adequação linguística perfeita, garantindo que a consulta nunca resulte em uma tela de "zero resultados" devido à incompletude dos dados.

Figura 6: Resultado da Busca por Função - Idioma Alternativo

Efeitos TRIZ

Idioma: Português

Problema: ☒ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito

Perspectiva Industrial: ☐ CNAE

Perspectiva Patentária: ☐ IPC

Ação: Desmembrar

Objeto: Sólido

Solução:

Buscar

124 Efeitos TRIZ Encontrados

- Ablação
- Ablação por laser
- Abrasão
- Acoustic Vibration

Este resultado foi exibido em Inglês, pois não foi encontrada uma versão no idioma selecionado.

Fonte: Elaborado pelo autor

Adicionalmente, para aumentar a Usabilidade e a eficácia do processo de busca, um mecanismo de validação de campos obrigatórios foi implementado na interface, conforme a Figura 7. Conforme ilustrado, ao tentar iniciar a recuperação da informação sem fornecer todos os parâmetros necessários para o modo de busca selecionado, o sistema exibe uma mensagem de erro ("Este campo é obrigatório"). Esta funcionalidade cumpre o princípio de Prevenção de Erros e Visibilidade do Status do Sistema (Nielsen, 1993), guiando o usuário a inserir os dados necessários. Ao fornecer orientações imediatas, o artefato reduz a taxa de falha na consulta e otimiza a eficiência do processo de RI, pois evita que o sistema processe buscas incompletas.

Figura 7: Mensagens de Erro para Buscas Incompletas

Efeitos TRIZ

Idioma: Português

Problema: ☒ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito

Perspectiva Industrial: ☐ CNAE

Perspectiva Patentária: ☐ IPC

Ação:

Objeto:

Solução:

Este campo é obrigatório.

Este campo é obrigatório.

Buscar

Agradecimentos:
Agradecemos à UFSCar e ao IBICT pelo apoio financeiro na realização deste projeto.

IBICT

Fonte: Elaborado pelo autor

Além do modo funcional (Função), o artefato permite ao usuário expressar sua necessidade informacional através de dois outros paradigmas de busca na seção Problema:

Parâmetro e Transformação. O modo Parâmetro foca a busca por "Operação" e "Parâmetro" a ser modificado (Ex.: Alterar Brilho), conforme mostra a Figura 8, permitindo ao usuário resolver conflitos técnicos baseados em atributos físicos ou mensuráveis. Já o modo Transformação, de acordo com a Figura 9, foca a busca por conversão de um estado de matéria ou energia para outro (Ex.: Transformar De Elétrica Para Química), auxiliando na identificação de efeitos que realizam conversões.

Figura 8: Busca por Parâmetro

Efeitos TRIZ

Idioma: Português

Problema: ☐ Funcão ☒ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito

Perspectiva Industrial: ☐ CNAE

Perspectiva Patentária: ☐ IPC

Operação: Alterar

Parâmetro: Brilho

Solução:

Buscar

114 Efeitos TRIZ Encontrados

- Absorção (física)
- Absorptive Filter
- Acousto-optic Effect
- Anéis de Newton

Fonte: Elaborado pelo autor

Figura 9: Busca por Transformação

Efeitos TRIZ

Idioma: Português

Problema: ☐ Funcão ☐ Parâmetro ☒ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito

Perspectiva Industrial: ☐ CNAE

Perspectiva Patentária: ☐ IPC

Transformar De: Elétrica

Para: Química

Solução:

Buscar

25 Efeitos TRIZ Encontrados

- Anodização
- Antifuse
- Arc Evaporation
- Bateria chumbo-ácido
- Bateria (eletricidade)

Fonte: Elaborado pelo autor

Embora possuam requisitos de entrada distintos — o modo Parâmetro utilizando o binômio Operação-Parâmetro e o modo Transformação utilizando o binômio De-Para —,

ambos compartilham a lógica de compatibilidade dinâmica implementada na busca por Função. O sistema utiliza o modelo de dados relacional (Seção 4.3) para garantir que a seleção do primeiro campo restrinja imediatamente as opções do segundo. Essa padronização funcional e a eliminação de combinações inválidas são importantes para a Usabilidade, oferecendo flexibilidade sem comprometer a eficiência da consulta. O resultado da recuperação segue o mesmo padrão de exibição em acordeão, permitindo a análise rápida e a mediação do conteúdo detalhado.

O modo Sugestão TRIZ (Efeito) é a estratégia de recuperação para usuários que já possuem um termo científico ou inventivo conhecido. Esta busca utiliza um campo de texto livre com sugestão em tempo real, como ilustrado na Figura 10, um recurso de Usabilidade projetado para acelerar a localização do efeito desejado e evitar erros de digitação. Ao selecionar o efeito, a estrutura de apresentação dos resultados é reversa à busca por problema, dedicando-se a detalhar o conhecimento sobre o princípio inventivo.

Figura 10: Busca por Efeito

Fonte: Elaborado pelo autor

A tela de resultado do efeito, conforme a Figura 11, apresenta o detalhamento completo (descrição, *links* IPC/CNAE), e lista os problemas específicos (pares Ação e Objeto) que o princípio é capaz de solucionar. Estes pares Ação-Objeto não são estáticos; cada combinação é apresentada com um ícone ativo (a lupa), que permite ao usuário executar instantaneamente uma nova consulta no modo Busca por Função. Este recurso é um exemplo prático de Reutilização (*Reusable*) e Interoperabilidade, pois transforma o resultado da busca em um novo ponto de partida, eliminando a necessidade de digitar novamente e promovendo a exploração contínua e eficiente do repositório.

Figura 11: Resultado da Busca por Efeito

Efeitos TRIZ

Idioma

Português

Problema

☐ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ

☐ Busca por IA ☒ Efeito

Perspectiva Industrial

☐ CNAE

Perspectiva Patentária

☐ IPC

Nome do Efeito

Ablação

Buscar

Ablação

Ablação é a remoção, vaporização, lascamento ou destruição por outros processos erosivos. Esse termo técnico é usado em astrofísica, biologia genética, cirurgia, engenharia e geologia para descrever fenômenos muito diferentes: Danos a espaçonave na reentrada atmosférica, deleção de genes, cirurgia a laser, tinta com anti-incrustantes, capas de proteção contra fogo e erosão durante glaciação.

IPC: G06F9/455

CNAE: 2829199

Ablação soluciona os problemas:

AÇÃO	OBJETO	
Absorver	Área	
Desmembrar	Sólido	
Mudança de fase	Sólido	
Limpar	Sólido	

Fonte: Elaborado pelo autor

O modo Perspectiva Industrial (CNAE) representa a estratégia de busca que utiliza vocabulário controlado para estabelecer a conexão direta entre o conhecimento inventivo e o contexto socioeconômico e produtivo brasileiro. Para otimizar a Usabilidade, a interface para este modo utiliza o mecanismo de autocompletar e sugestão em tempo real, conforme ilustrado na Figura 13, que auxilia o usuário na inserção do código ou da descrição da atividade econômica (CNAE), limitando as opções e permitindo o uso exclusivo de termos válidos e mapeados na base de dados.

Figura 12: Busca por CNAE

Efeitos TRIZ

Idioma dos Resultados

Português

Problema

☐ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ

☐ Busca por IA ☐ Efeito ☒ CNAE

Perspectiva Industrial

☐ CNAE

Perspectiva Patentária

☐ IPC

Atividade Industrial ou Código CNAE

agênci

Buscar

6391-7/00 - AGÊNCIAS DE NOTÍCIAS

6434-4/00 - AGÊNCIAS DE FOMENTO

7311-4/00 - AGÊNCIAS DE PUBLICIDADE

7911-2/00 - AGÊNCIAS DE VIAGENS

9609-2/02 - AGÊNCIAS MATRIMONIAIS

Agri

Agri

Fonte: Elaborado pelo autor

Ao selecionar um código CNAE, o sistema recupera todos os efeitos TRIZ que lhes são associados, como ilustra a Figura 13. O formato de exibição dos resultados é padronizado e visa à Reutilização (*Reusable*) da informação: os detalhes de cada efeito são apresentados em acordeão, acompanhados de recursos visuais, quando houver, (animações/GIFs) e dos vínculos

de classificação externos (WIPO, Espacenet e IBGE). A tela de resultado lista os problemas específicos (pares Ação → Objeto) que o efeito recuperado soluciona, como mostra a Figura 13. Essa inversão da lógica (do Efeito para o Problema) transforma o resultado da busca em um guia prático de aplicação, facilitando o uso do conhecimento para a resolução de tarefas inventivas e reforçando a função do artefato como um nó de informação tecnológica.

Figura 13: Resultado da Busca por CNAE

Efeitos TRIZ

Idioma dos Resultados: Português

Problema: ☐ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito ☒ CNAE ☐ IPC

Perspectiva Industrial: ☐ CNAE ☐ IPC

Atividade Industrial ou Código CNAE: 6434400 - AGÊNCIAS DE FOMENTO

Efeito Matteucci (6)

Efeito Matteucci é um dos efeitos magnetomecânicos, que é termodinamicamente inversa a Efeito Wiedemann. Este efeito foi descrito por Carlo Matteucci em 1858.

IPC: A61P35/00 **WIPO**

Efeito Matteucci soluciona os problemas:

Problema	Tipo
Transformação: Cinética → Magnético	Efeito
Função: Produzir → Área	Efeito
Função: Girar → Área	Efeito
Parâmetro: Medir → Força	Efeito
Parâmetro: Medir → Orientação	Efeito
Transformação: Mecânica → Magnético	Efeito

Fonte: Elaborado pelo autor

De modo complementar à estratégia industrial, o modo Perspectiva Patentária estabelece a conexão do conhecimento inventivo com o contexto tecnológico formalizado pelos documentos de patentes. Conforme ilustrado na Figura 14, ao selecionar a opção IPC, a interface adapta-se dinamicamente e habilita um subnível para a escolha do "Modo de Busca Patentária". Nesta etapa, o sistema oferece flexibilidade ao permitir que a consulta seja realizada por meio de duas abordagens distintas: "Por Código IPC" ou "Por Palavra-chave WIPO".

Figura 14: Opções de Busca da Perspectiva Patentária

Efeitos TRIZ

Idioma: Português

Problema: ☐ Função ☐ Parâmetro ☐ Transformação

Sugestão TRIZ: ☐ Busca por IA ☐ Efeito ☒ CNAE ☐ IPC

Perspectiva Industrial: ☐ CNAE ☐ IPC

Modo de Busca Patentária: ☒ Por Código IPC ☐ Por Palavra-chave WIPO

Código IPC: C02F1/52

Agradecimentos:
Agradecemos à UFSCar e ao IBICT pelo apoio financeiro na realização deste projeto.

Fonte: Elaborado pelo autor

A busca "Por Código IPC", na Figura 15, possibilita a inserção direta da classificação (como, por exemplo, "C09K8/52"), sendo ideal para pesquisas estruturadas em que o usuário já possui a referência da patente e exibe os resultados de forma semelhante à Figura 13, das buscas por CNAE.

Figura 15: Busca por Código IPC

A interface de busca "Efeitos TRIZ" apresenta uma barra de pesquisa com o seguinte layout:

- Idioma:** Dropdown menu com "Português" selecionado.
- Problema:** Três opções de radio button: "Função", "Parâmetro" e "Transformação".
- Sugestão TRIZ:** Três opções de radio button: "Busca por IA", "Efeito" e "CNAE".
- Perspectiva Industrial:** Uma opção de radio button.
- Perspectiva Patentária:** Duas opções de radio button, com "IPC" selecionada.
- Modo de Busca Patentária:** Duas opções de radio button, com "Por Código IPC" selecionada e "Por Palavra-chave WIPO" desselecionada.
- Código IPC:** Campo de texto com o valor "C09K8/52" inserido.
- Botão:** Botão azul "Buscar" no canto inferior direito.
- Resultado:** Abaixo da barra, há uma seção com o texto "Agglomeration (1)" e uma seta para baixo.

Fonte: Elaborado pelo autor

A opção "Por Palavra-chave WIPO", conforme a Figura 16, materializa na interface a integração do vocabulário do índice *Catchwords*. Esta alternativa viabiliza a recuperação da informação patentária através de termos em linguagem natural, mitigando necessidade de o usuário conhecer previamente os códigos hierárquicos do sistema internacional de patentes. A adoção deste índice introduz uma especificidade linguística ao sistema. Como o vocabulário oficial de *catchwords* da WIPO é disponibilizado apenas em inglês e francês, a interface solicita que o usuário defina em qual destes dois idiomas a consulta será realizada, independentemente do idioma geral do sistema. Diferentemente dos modos baseados em códigos exatos, este campo opera como uma busca textual livre sem o mecanismo de autocompletar, exigindo a inserção do termo para que o sistema processe no banco de dados.

Ao realizar a pesquisa por palavra-chave, a interface apresenta primeiramente a hierarquia completa do termo localizado, contextualizando semanticamente a busca do usuário. Em seguida, o sistema lista os códigos IPC correspondentes àquela palavra-chave e, sob cada código, elenca os respectivos efeitos TRIZ mapeados. A visualização detalha, de forma direta, quais pares de Tarefa e Alvo (ou parâmetros de Transformação) e seus respectivos tipos respondem àquele problema específico, acompanhados de seu grau de confiabilidade (*Score*).

Figura 16: Busca por Palavra-chave WIPO

The screenshot shows the 'Efeitos TRIZ' web application. At the top, there are tabs for 'Idioma' (set to Português), 'Problema' (with radio buttons for Função, Parâmetro, and Transformação), 'Sugestão TRIZ' (with radio buttons for Busca por IA, Efeito, CNAE, and IPC, where IPC is selected), 'Perspectiva Industrial', and 'Perspectiva Patentária'. Below these, the 'Modo de Busca Patentária' is set to 'Por Palavra-chave WIPO'. A blue box contains a warning: 'Atenção: Os índices da WIPO estão disponíveis apenas nos idiomas abaixo. Selecione para sua busca:' with radio buttons for English (selected) and Français. The 'Palavra-chave de Busca' field contains 'abrasive', and a 'Buscar' button is to its right. The results section, titled 'PALAVRA-CHAVE ABRASIVE(S) > ABRASIVE(S) materials', shows 'Patentes e Efeitos TRIZ Mapeados' with a filter for 'IPC: C09K3/14 WIPO'. A result for 'Cohesion' is shown with the effect 'Effect' and transformation 'Transformation: Chemical → Mechanical', with a 'Score: 3'.

Fonte: Elaborado pelo autor

Avançando nas estratégias de mediação, o sistema apresenta a modalidade de "Busca por IA", localizada sob o eixo "Sugestão TRIZ", conforme a Figura 17. Esta funcionalidade introduz uma interface conversacional (*chatbot*) que permite ao usuário formular suas necessidades de informação tecnológica em linguagem natural. Ao inserir uma pergunta, o sistema aciona a arquitetura de RAG. O sistema demonstrou ser possível a recuperação de informação ao processar cada consulta em duas camadas: a matemática (via banco de dados vetorial *Pinecone*) e a lógica (via *framework* Laravel). A consulta textual é processada semanticamente, localizando os registros com maior proximidade matemática ao sentido da pergunta do usuário. O fluxo conversacional é pautado no idioma selecionado pelo usuário.

Figura 17: Busca Semântica com IA

The screenshot shows the 'Efeitos TRIZ' web application with the 'Sugestão TRIZ' tab set to 'Busca por IA'. The 'Idioma' is 'Português'. A search box contains the question 'Como diminuir atrito?' with a 'Buscar' button. Below the search box, a section titled 'ANÁLISE DE IA SEMÂNTICA' provides a detailed answer: 'Para diminuir o atrito, as seguintes abordagens são indicadas: Lubrificação: A aplicação de substâncias lubrificantes reduz o atrito entre superfícies em contato. Rolamento: A utilização de rolamentos substitui o atrito de deslizamento por atrito de rolamento, que é significativamente menor. Amortecimento de Coulomb: Este efeito, quando diminuído, contribui para a redução do atrito. Revestimento: A aplicação de revestimentos pode modificar as propriedades superficiais, diminuindo o atrito. Essas técnicas estão associadas a diversas áreas, incluindo o COMÉRCIO; REPARAÇÃO DE VEÍCULOS AUTOMOTORES E MOTOCICLETAS, INDÚSTRIAS DE TRANSFORMAÇÃO e SAÚDE HUMANA E SERVIÇOS SOCIAIS.' At the bottom of this section is a link 'VER FONTES TÉCNICAS CONSULTADAS (100 EFEITOS)'. Below the answer is another search box with the text 'Como diminuir o atrito...' and a 'Buscar' button.

Fonte: Elaborado pelo autor

Como ilustrado na Figura 18, logo abaixo da resposta fornecida, é possível expandir e visualizar quais efeitos foram utilizados para formular a resposta. Essa abordagem ancora o modelo generativo no contexto técnico. Diretrizes no *prompt* orientam a IA a restringir suas respostas ao conteúdo presente, evitando a geração de informações externas à base.

Figura 18: Fontes Técnicas do RAG

The screenshot displays the 'Efeitos TRIZ' (TRIZ Effects) interface. At the top, there are tabs for 'Idioma' (Language) set to 'Português', 'Problema' (Problem) with options 'Função', 'Parâmetro', 'Transformação', and 'Sugestão TRIZ' (TRIZ Suggestion) with 'Busca por IA' (AI Search) selected. There are also tabs for 'Perspectiva Industrial' (Industrial Perspective) with 'CNAE' selected, and 'Perspectiva Patentária' (Patent Perspective) with 'IPC' selected. Below these, a list of search results is shown, each with an ID, a description in Portuguese, and a match percentage. The results include:

- ID: 21564, Match: 78,9%
- ID: 21565, Match: 78,9%
- ID: 13861, Match: 77,3%
- ID: 13862, Match: 77,3%
- ID: 13863, Match: 77,3%
- ID: 1489, Match: 76,3%

At the bottom, there is a search bar with the text 'Como diminuir o atrito...' and a 'Buscar' (Search) button.

Fonte: Elaborado pelo autor

Apesar dos avanços proporcionados pela interface conversacional, a avaliação empírica do artefato revelou certas limitações inerentes à estratégia de vetorização e de formulação de contexto adotadas. Observou-se que o método de concatenação dos dados para a alimentação do banco vetorial gerou um viés semântico na recuperação da informação. Por possuírem um volume textual significativamente maior, as descrições qualitativas dos efeitos TRIZ exerceram um peso desproporcional no cálculo de similaridade matemática no espaço vetorial. Como consequência, consultas direcionadas especificamente à localização de vocabulários controlados, como os códigos IPC ou CNAE, apresentaram um nível de similaridade baixo, tornando a busca semântica por esses eixos pouco efetiva. Na prática, a recuperação baseada em *embeddings* tendeu a privilegiar o conteúdo descritivo em detrimento da precisão classificatória oferecida pelas outras perspectivas do sistema.

Outro aspecto relevante diz respeito ao compromisso entre a mitigação de alucinações e a fluidez conversacional. A imposição de diretrizes estritas no *prompt* de sistema, aliada à necessidade processar um contexto espesso contendo o mapeamento simultâneo nos quatro idiomas, demonstrou-se eficaz para ancorar as respostas e preservar a integridade técnica da base. Contudo, essa abordagem de ancoragem rígida possibilitou o modelo generativo adotar

um comportamento próximo ao literal. Embora essa literalidade seja um requisito fundamental para assegurar a correspondência em domínios como o de patentes e o industrial, ela restringe a flexibilidade inata do *chatbot*. Como resultado, o modelo apresenta dificuldades em sustentar diálogos mais amplos ou conduzir abstrações exploratórias que fujam da literalidade do contexto, limitando parcialmente a experiência de conversação aberta em favor da precisão documental.

Para a conclusão da análise funcional do artefato, além das particularidades do modelo generativo, é importante validar a funcionalidade multilíngue em sua totalidade, conforme o requisito central do projeto. A interface foi projetada com um mecanismo de internacionalização (i18n), demonstrado pelo seletor "Idioma" no cabeçalho. Ao alternar o idioma, o sistema executa um processo em duas camadas para garantir a coerência:

- **Mapeamento da Interface (Camada Estática):** Um arquivo de mapeamento de elementos textuais é acionado para trocar o idioma de todos os elementos textuais estáticos da interface (rótulos de campos, títulos de seção). O que permite que o usuário interaja com o sistema em seu idioma de preferência.
- **Recuperação do Conteúdo (Camada Dinâmica):** Simultaneamente, a consulta ao banco de dados utiliza a chave do idioma selecionado para acessar os dados armazenados em JSONB (Seção 4.3), garantindo que os nomes dos efeitos, tarefas e alvos retornados também estejam no idioma correto.
- **Processamento Semântico (Camada Generativa):** Quanto à inteligência artificial, o sistema transfere o parâmetro de idioma para as diretrizes de processamento do modelo generativo. Assim, permite que a extração de entidades, a síntese técnica via RAG e o diálogo do *chatbot* operem estritamente na mesma língua da interface, preservando a padronização terminológica.

Esta arquitetura multilíngue integrada promove uma experiência na qual a interface, o conteúdo estruturado da base e as interações baseadas em IA convergem para o idioma do usuário.

Em suma, o desenvolvimento do artefato resultou em um sistema funcional, acessível para consulta pública no endereço eletrônico <https://trizeffectdatabase.quoniam.info/>. A disponibilização via HTTPS (*Hypertext Transfer Protocol Secure*) reforça o compromisso do projeto com a segurança e a integridade dos dados trocados, estabelecendo um canal confiável de comunicação e validação para a comunidade científica e o setor industrial.

5 DISCUSSÃO DOS RESULTADOS

A principal conclusão empírica do trabalho reside na operacionalização do processo de tradução da informação (Cesarino, 1985), que viabilizou a conversão do estoque estático da planilha original em um fluxo dinâmico e utilizável. A arquitetura implementada — baseada na seção (4.3) com a solução multilíngue JSONB e nos princípios de Usabilidade (IHC) — resolveu a barreira de transferência (Barreto, 1996) identificada na fase diagnóstica. O artefato não apenas disponibiliza os dados, mas o faz com técnicas de Recuperação da Informação (RI), sustentada pela arquitetura de busca dinâmica e pela Usabilidade que minimiza a taxa de falha na consulta (Krug, 2006).

A aderência aos princípios A, I e R reforça a premissa de que o investimento em um processo de Organização e Representação da Informação ativo é indispensável para transformar dados abertos em ativos estratégicos. O valor agregado do sistema se concentra nas estratégias de recuperação contextual (CNAE e IPC), que estabelecem um novo ponto de contato entre o conhecimento científico abstrato (TRIZ) e os contextos econômicos e tecnológicos do Brasil. No domínio patentário, a integração do índice *Catchwords* ampliou as rotas de acesso ao viabilizar consultas em linguagem natural, embora exija do usuário a adaptação a uma especificidade linguística (inglês e francês) ditada pela organização fonte (WIPO). Desta forma, o artefato desempenha o papel de mediador e catalisador para a aplicação da informação tecnológica no campo da inovação e prospecção.

Adicionalmente, a incorporação de LLM e RAG representou uma evolução na etapa de busca do sistema, viabilizando a estratégia semântica conversacional. Contudo, a avaliação empírica evidenciou compromissos técnicos inerentes a essa adoção. Observou-se um viés no processamento vetorial, que tendeu a privilegiar o volume textual das descrições dos efeitos em detrimento de buscas direcionadas por classificações exatas. Ademais, a necessidade de ancoragem estrita no contexto recuperado para mitigar alucinações (Lewis et al., 2020) forçou a inteligência artificial a adotar um comportamento próximo do literal. Isso aponta que, embora a tecnologia generativa enriqueça a recuperação temática, ela propicia limitações à fluidez de diálogos exploratórios mais amplos em cenários que exigem precisão documental.

Em última análise, este trabalho demonstra a Ciência da Informação disciplina integradora capaz de orquestrar técnicas computacionais, princípios de Usabilidade e estratégias de Recuperação da Informação para operacionalizar os ideais da Ciência Aberta. A contribuição fundamental reside na proposição de um modelo de mediação técnica que viabiliza a conversão de um estoque de dados brutos em um fluxo de conhecimento organizado e

acessível. Ao estruturar um repositório especializado, fundamentado no conteúdo técnico extraído de documentos de patentes, o artefato favorece a interoperabilidade e o potencial de reuso desse conhecimento (FAIR). Desta forma, a informação tecnológica constitui-se como um insumo estratégico, apto a subsidiar a inovação e o desenvolvimento socioeconômico.

6 CONSIDERAÇÕES FINAIS

A presente pesquisa, caracterizada como empírica-descritiva-exploratória e aplicada, teve como objetivo geral propor um modelo de recuperação da informação para a Base de efeitos TRIZ Multilíngue, visando mitigar as barreiras de transferência impostas pelo seu formato original em planilha. O ponto de partida foi o diagnóstico de que o formato planilhado, embora alinhado à racionalidade técnica de produção de estoque, criava obstáculos para o fluxo e a transferência do conhecimento, um desafio fundamental na Ciência da Informação.

Para alcançar esse objetivo principal, a pesquisa seguiu um percurso metodológico rigoroso. Primeiramente, realizou-se uma análise da estrutura informacional da base de efeitos TRIZ de Zaniro e Quoniam (2024), que identificou a riqueza de 47 colunas, abrangendo metadados descritivo-semânticos, ligações externas (IPC e CNAE), ampliadas pela integração textual do vocabulário de *Catchwords* da WIPO, e elementos multilíngues, confirmando a complexidade que demandava uma reestruturação sistêmica. Em paralelo, foram exploradas bases de efeitos TRIZ correlatas (Aulive e Oxford Effects Database), o que forneceu *benchmarking* para as estratégias de busca e Usabilidade, validando a necessidade de flexibilidade de consulta e, sobretudo, retificando a lacuna de acessibilidade linguística que o trabalho buscou endereçar.

O ponto central da pesquisa reside na proposta do artefato de mediação. A conformidade do sistema foi avaliada por meio dos Princípios FAIR (*Findable, Accessible, Interoperable, Reusable*), os quais serviram como requisitos arquiteturais e métrica de sucesso. A análise demonstrou que a estrutura relacional, o uso de protocolos abertos (HTTP/HTTPS) e a vinculação a identificadores externos padronizados permitiram a aderência aos pilares da Acessibilidade, Interoperabilidade e Reutilização. Em complemento à arquitetura relacional, a evolução do artefato para um modelo híbrido — ao integrar uma camada conversacional baseada em IA e RAG — representou um avanço na flexibilização da Recuperação da Informação.

Esses resultados sugerem, de forma empírica, que o investimento em um processo ativo de Organização e Representação da Informação é um fator relevante para transformar dados abertos de um estoque estático em um fluxo dinâmico e utilizável. O impacto direto do trabalho

consiste na contribuição para redução da barreira de transferência, posicionando a base a um mediador entre o conhecimento inventivo (TRIZ) e os contextos econômicos e tecnológico, como sugere a Ciência da Informação.

Como desdobramento das etapas de investigação e desenvolvimento do artefato, os resultados obtidos neste projeto contribuíram para a literatura da área. Destacam-se as publicações apresentadas no XXV Encontro Nacional de Pesquisa e Pós-graduação em Ciência da Informação (ENANCIB) em 2025, que abordaram a utilização de *Linked Data* em repositórios de informação tecnológica (ERBA; ZANIRO; QUONIAM, 2025) e a relação entre Ciência da Informação, patentes e comunidades indígenas (ERBA et al., 2025). Adicionalmente, as discussões sobre a análise e organização da informação tecnológica (ERBA; ZANIRO, 2024a) e suas contribuições para a conservação de biomas brasileiros (ERBA; ZANIRO, 2024b) foram publicadas, respectivamente, no V Congresso Brasileiro Interdisciplinar em Ciência e Tecnologia (COBICET) e no 14º Congresso Científico da Semana Nacional de Ciência e Tecnologia do IFSP (CONCISTEC) em 2024. Estas publicações demonstram a aderência temática da pesquisa no âmbito acadêmico, reafirmando o potencial da informação tecnológica e dos dados estruturados para a inovação e o desenvolvimento social.

REFERÊNCIAS

ALTSHULLER, G. S. **The innovation algorithm: TRIZ, systematic innovation & technical creativity**. Tradução de: Lev Shulyak; Steven Rodman. Worcester, MA: Technical Innovation Center, 1999.

ALVARES, L. M. A. de R.; ITABORAHY, A. L. C. (org.). **Os múltiplos cenários da informação tecnológica no Brasil do século XXI**. Rio de Janeiro: IBICT, 2021. 474 p.

ARAÚJO, V. M. R. H. Conceitos básicos da informação tecnológica. **In: Seminário Integração das agências de fomento para informação tecnológica**, 1993, Brasília, **Anais [...]**. Brasília: IBICT, 1993.

ARAÚJO, R. F. *et al.* Ciência aberta na perspectiva de especialistas brasileiros: proposta de taxonomia. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, v. 26, p. 1-27, 2021.

AULIVE. **Aulive**. Disponível em: <https://www.aulive.com/>. Acesso em: 20 set. 2025.

BARRETO, A. de A. A eficiência técnica e econômica e a viabilidade de produtos e serviços de informação. **Ciência da informação**, v. 25, n. 3, p. 405-414, 1996.

BARROS, W. B. G. *et al.* Patente como fonte de informação tecnológica: utilização de documentos de patente em domínio público. **In: SIMPÓSIO INTERNACIONAL DE GESTÃO DE PROJETOS**, 3.; **SIMPÓSIO INTERNACIONAL DE INOVAÇÃO E SUSTENTABILIDADE**, 2., 2014, São Paulo. **Anais [...]**. São Paulo: SINGEP, 2014. p. 1-10. ISSN 2317-8302.

BUCKLAND, M. K. Information as thing. **Journal of the American Society for Information Science**, v. 42, n. 5, p. 351-360, 1991.

BARROSO, W.; QUONIAM, L.; PACHECO, E. Patents as technological information in Latin America. **World Patent Information**, v. 31, n. 3, p.207–215, 2009. DOI: <https://doi.org/10.1016/j.wpi.2008.11.006>.

CAFEZEIRO, I.; COSTA, L. C.; KUBRUSLY, R. S. Ciência da Computação, Ciência da Informação, Sistemas de Informação: Uma reflexão sobre o papel da informação e da interdisciplinaridade na configuração das tecnologias e das ciências. **Perspectivas em Ciência da Informação**, v. 21, p. 111-133, 2016.

CAPURRO, R. Past, present, and future of the concept of information. **TripleC: communication, capitalism & critique**. open access journal for a global sustainable information society, v. 7, n. 2, p. 125-141, 2009.

CARVALHO, M. A. **Inovação em produtos: IDEATRIZ, uma aplicação da TRIZ / Inovação Sistemática na ideação de produtos**. 2. ed. São Paulo: Blucher, 2017.

CASTRO, A. C.; JANNUZZI, C. A. S. C.; MATTOS, F. A. M. Produção e disseminação de informação tecnológica: a atuação da Inova - Agência de Inovação da UNICAMP. **TransInformação**, Campinas, v. 19, n. 3, p. 265-277, set./dez. 2007.

CESARINO, M. A. N. Sistemas de recuperação da informação. **Revista da Escola de Biblioteconomia da UFMG**, v. 14, n. 2, 1985.

CORMEN, Thomas H. *et al.* Algoritmos: teoria e prática. **Editora Campus**, v. 2, p. 296, 2002.

CREATIVE COMMONS BRASIL. **As licenças**. [S. l.], 2025. Disponível em: <https://br.creativecommons.net/licencas/>. Acesso em: 10 dez. 2025.

CREATIVE COMMONS. **Attribution-ShareAlike 4.0 International (CC BY-SA 4.0)**. Disponível em: <https://creativecommons.org/licenses/by-sa/4.0/>. Acesso em: 22 mai. 2025.

DE ALBUQUERQUE, Ana Cristina. Em foco a classificação: abordagens conceituais na Arquivologia, Biblioteconomia e Museologia. **Encontros Bibli: revista eletrônica de biblioteconomia e ciência da informação**, v. 20, n. 43, p. 20-46, 2015.

ERBA, A. G.; ZANIRO, D. L. Análise e organização da informação tecnológica com o apoio da Ciência da Computação. In: V CONGRESSO BRASILEIRO INTERDISCIPLINAR EM CIÊNCIA E TECNOLOGIA (V COBICET), 2024. **Anais do V COBICET**, 2024a.

ERBA, A. G.; ZANIRO, D. L. Contribuições da informação tecnológica para a conservação e revitalização dos biomas brasileiros: Análise das Respostas Técnicas do SBRT. In: 14º CONGRESSO CIENTÍFICO DA SEMANA NACIONAL DE CIÊNCIA E TECNOLOGIA DO IFSP (CONCISTEC 2024), 2024, Bragança Paulista - SP. **Anais do CONCISTEC 2024**, 2024b.

ERBA, A. G.; ZANIRO, D. L.; QUONIAM, L. Investigando a utilização de Linked Data em um repositório de informação tecnológica. In: XXV ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (XXV ENANCIB), 2025, Rio de Janeiro - RJ. **Anais do XXV ENANCIB**, 2025.

ERBA, A. G.; ZANIRO, D. L.; QUONIAM, L.; DAL'EVEDOVE, P. R. Ciência da Informação, patentes e comunidades indígenas. In: XXV ENCONTRO NACIONAL DE PESQUISA E PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO (XXV ENANCIB), 2025, Rio de Janeiro - RJ. **Anais do XXV ENANCIB**, 2025.

EUROPEAN PATENT OFFICE. **Espacenet**: patent search. [S. l.], 2025. Disponível em: <https://worldwide.espacenet.com/>. Acesso em: 05 dez. 2025.

FERNEDA, E. **Recuperação de informação: análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. 137 f. 2003. Tese de Doutorado. Tese (Doutorado em Ciência da Informação e Documentação) - Universidade de São Paulo. São Paulo.

GARCIA, J. C. R.; CHACON, F. J. F. O ensino da Classificação Internacional de Patentes (CIP) nos cursos de Biblioteconomia brasileiros. **Informação & Informação**, Londrina, v. 13, n. 2, p. 15-33, jul./dez. 2008.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4. ed. São Paulo: Atlas, 2002

GOODFELLOW, I. *et al.* **Deep learning**. Cambridge: MIT press, 2016.

GOOGLE. Gemini: a family of highly capable multimodal models. **Google DeepMind**, 2023. Disponível em: https://storage.googleapis.com/deepmind-media/gemini/gemini_1_report.pdf. Acesso em: 16 mar. 2026.

HENNING, P. *et al.* Desmistificando os princípios FAIR: conceitos, métricas, tecnologias e aplicações inseridas no ecossistema dos dados FAIR. **Tendências da Pesquisa Brasileira em Ciência da Informação**, v. 11, n. 1, 2018.

HENNING, P. C. *et al* GO FAIR e os princípios FAIR: o que representam para a expansão dos dados de pesquisa no âmbito da Ciência Aberta. **Em Questão**, Porto Alegre, v. 25, n. 2, p. 389-412, 2019. DOI: 10.19132/1808-5245252.389-412. Acesso em: 10 fev. 2024.

HINNOSAAR, M.; HINNOSAAR, T.; KUMMER, M.; SLIVKO, O. Wikipédia matters. *Journal of Economics & Management Strategy*, v. 32, n. 3, p. 657-669, 2023. Disponível em: <https://doi.org/10.1111/jems.12421>. Acesso em: 8 maio 2025.

INSTITUTO BRASILEIRO DE GEOGRAFIA E ESTATÍSTICA (IBGE). **Classificação Nacional de Atividades Econômicas - CNAE**. 2.1 ed. 2022. Disponível em: <https://concla.ibge.gov.br/classificacoes/cnae.html>. Acesso em: 22 nov. 2024.

IBGE. **Classificação Nacional de Atividades Econômicas - CNAE**. Rio de Janeiro: IBGE, 2025. Disponível em: <https://cnae.ibge.gov.br/>. Acesso em: 12 maio 2025.

JURAFSKY, D.; MARTIN, J. H. **Speech and language processing**. Vol. 3. 2014. Disponível em: <https://web.stanford.edu/~jurafsky/slp3/>.

KRISNAWATI, L. D.; MAHASTAMA, A. W. The Scenario of Wikipédia Usage as a supporting source for under-resource language researches. **PRISMA.COM**, Porto, n. 40, p. 34-44, 2019.

KRUG, Steve. **Não me faça pensar: uma abordagem de bom senso à Usabilidade na Web**. 2. ed. Rio de Janeiro: Alta Books, 2006.

LEWIS, P. *et al.* Retrieval-augmented generation for knowledge-intensive NLP tasks. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS**, 33., 2020. Anais [...], 2020. p. 9459–9474. Disponível em: <https://proceedings.neurips.cc/paper/2020/file/6b493230205f780e1bc26945df7481e5-Paper.pdf>. Acesso em: 21 mar. 2026.

LI, S. P.; YU, K. M.; YEUNG, Y. C.; KEUNG, K. L. "Patent review and novel design of vehicle classification system with TRIZ" **World Patent Information**, v. 71, 2022. DOI: <https://doi.org/10.1016/j.wpi.2022.102155>

LIN, Dawei *et al.* The TRUST Principles for digital repositories. **Scientific Data**, v. 7, n. 144, 2020. DOI: 10.1038/s41597-020-0486-7.

MARTIN, A. Brewing free beer: using ideality to develop a ‘free-to-use’ TRIZ effects database. **Procedia Engineering**, v. 131, p. 312-326, 2015.

MOEHRLE, M. G. How combinations of TRIZ tools are used in companies—results of a cluster analysis. **R&D Management**, v. 35, n. 3, p. 285-296, 2005.

NIELSEN, Jakob. **Usability engineering**. Morgan Kaufmann, 1994.

OLIVEIRA JUNIOR, A. M.; ALMEIDA, J. J. M. Análise das inter-relações das patentes das universidades sob a perspectiva de mercado. **Navus: Revista de Gestão e Tecnologia**, Florianópolis, v. 9, n. 4, p. 139-160, out./dez. 2019.

OXFORD. **Effects Database**. Disponível em: <https://wbam2244.dns-systems.net/EDB/>. Acesso em: 20 set. 2025.

PARANAGUÁ, P.; REIS, R. **Patentes e criações industriais**. Rio de Janeiro: Editora FGV, 2009. (Série FGV Jurídica).

PIRES, E. A.; RIBEIRO, N. M.; QUINTELLA, C. M. Sistemas de busca de patentes: análise comparativa entre Espacenet, Patentscope, Google Patents, Lens, Derwent Innovation Index e Orbit Intelligence. **Cadernos de Prospeção**, Salvador, v. 13, n. 1, p. 1-16, 2020. DOI: 10.9771/cp.v13i1.35147.

PISCOPO, A. *et al.* What do Wikidata and Wikipédia Have in Common? An Analysis of their Use of External References. **Proceedings of the 13th International Symposium on Open Collaboration (OpenSym '17)**, Galway, Ireland, p. 1-10, ago. 2017. DOI: <https://doi.org/10.1145/3125433.3125445>.

PRODANOV, C. C. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico**. Novo Hamburgo: Feevale, 2013. 276 p.

RUSSELL, S.; NORVIG, P. Artificial Intelligence: A modern approach. **Prentice-Hall**, Englewood Cliffs, v. 25, n. 27, p. 79-80, 1995.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em ciência da informação**, v. 1, n. 1, 1996.

SARACEVIC, T. Relevance: A review of and a framework for the thinking on the notion in information science. **Journal of the American Society for information science**, v. 26, n. 6, p. 321-343, 1975.

SARACEVIC, T. Interdisciplinary nature of information science. **Ciência da informação**, v. 24, n. 1, p. 36-41, 1995.

SARACEVIC, T. Information Science. **Journal of the American Society for Information Science**, v. 50, n. 12, p. 1051-1063, 1999.

SERVIÇO NACIONAL DE APRENDIZAGEM INDUSTRIAL. **Departamento Nacional. Classificação das ações do SENAI: termos e conceitos dos serviços técnicos e tecnológicos**. Brasília: SENAI/DN, 2002. 27 p.

SILVA, F. C. C.; SILVEIRA, L. O ecossistema da Ciência Aberta. **TransInformação**, Campinas, v. 31, e190001, 2019. Disponível em: <http://dx.doi.org/10.1590/2318-0889201931e190001>. Acesso em: 04 jan. 2025.

SILVA, J. L. C. Percepções conceituais sobre mediação da informação. InCID: **Revista de Ciência da Informação e Documentação**, v. 6, n. 1, p. 93-108, 2015.

SOUZA, R.R. Sistemas de recuperação de informações e mecanismos de busca na *Web*: panorama atual e tendências. **Perspectivas em Ciência da Informação**, v. 11, n. 2, p. 161-173, maio/ago. 2006.

VASWANI, A. *et al.* Attention is all you need. In: **ADVANCES IN NEURAL INFORMATION PROCESSING SYSTEMS**, 30., 2017. *Anais [...]*, 2017. Disponível em: https://proceedings.neurips.cc/paper_files/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf. Acesso em: 21 mar. 2026.

WILKINSON, M. D. *et al.* The FAIR Guiding Principles for scientific data management and stewardship. **Scientific Data**, v. 3, p. 160018, 2016. DOI: 10.1038/sdata.2016.18. Disponível em: <https://www.nature.com/articles/sdata201618>. Acesso em: 03 dez. 2024.

WIPO. **WIPO Guide to Using Patent Information**. 2022. WIPO Publication nº L 434. Disponível em: <https://www.wipo.int/publications/en/details.jsp?id=4615>. Acesso em: 8 maio 2025.

WIPO. IPC. **International Patent Classification (IPC)**. 2024. Disponível em: <https://www.wipo.int/classifications/ipc/en/>. Acesso em: 23 de ago. 2025.

ZANIRO, D. L.; QUONIAM, L. **Base de dados de efeitos TRIZ multilíngue ligada com IPC e CNAE**. 24 maio 2024. Disponível em: <https://repositorio.ufscar.br/handle/ufscar/20138>. Acesso em: 07 nov. 2024. Coleção Dados de Pesquisa UFSCar.

ZANIRO, D. L.; QUONIAM, L.; SOUZA, M. G. de; CARVALHO SEGUNDO, W. L. R. de. Towards an open TRIZ multilingual database. In: INTERNATIONAL CONFERENCE ON OPEN REPOSITORIES (OR2024), 19., 2024, Göteborg. **Proceedings [...]**. Göteborg: [s.n.], 2024. Disponível em: https://www.researchgate.net/publication/381482544_Towards_an_open_TRIZ_multilingual_database. Acesso em: 20 nov. 2024.