

UNIVERSIDADE DE SÃO PAULO  
ESCOLA DE COMUNICAÇÕES E ARTES  
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

LUÍS ROBERTO MOMBERG ALBANO

**Modelo para recuperação de informações em ambientes digitais:**  
uma proposta baseada no contexto de uso

São Paulo  
2023

LUÍS ROBERTO MOMBERG ALBANO

**Modelo para recuperação de informações em ambientes digitais:  
uma proposta baseada no contexto de uso**

**Versão Corrigida**

**(Versão original disponível na Biblioteca da Escola de Comunicações e Artes  
da Universidade de São Paulo)**

Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação da Universidade de São Paulo para a obtenção do título de Doutor em Ciência da Informação.

Área de Concentração: Cultura e Informação.

Orientador: Prof. Dr. Marcelo dos Santos.

São Paulo  
2023

Autorizo a reprodução e divulgação total ou parcial deste trabalho, por qualquer meio convencional ou eletrônico, para fins de estudo e pesquisa, desde que citada a fonte.

Catálogo na Publicação  
Serviço de Biblioteca e Documentação  
Escola de Comunicações e Artes da Universidade de São Paulo  
Dados inseridos pelo autor

---

Albano, Luis Roberto Momberg  
Modelo para recuperação de informações em ambientes digitais: uma proposta baseada no contexto de uso / Luis Roberto Momberg Albano; orientador, Marcelo dos Santos. - São Paulo, 2023. - 239 p.: il.

Tese (Doutorado) - Programa de Pós-Graduação em Ciência da Informação / Escola de Comunicações e Artes / Universidade de São Paulo.

Bibliografia  
Versão corrigida

1. Sistema de Recuperação de informação. 2. Contexto de uso. 3. Ambientes digitais. I. Santos, Marcelo dos. II. Título.

CDD 21.ed. - 020

---

Elaborado por Alessandra Vieira Canholi Maldonado - CRB-8/6194

ALBANO, Luís Roberto Momberg. **Modelo para recuperação de informações em ambientes digitais**: uma proposta baseada no contexto de uso. 2023. Doutorado (Ciência da Informação) – Universidade de São Paulo - USP, São Paulo, SP, 2023.

Aprovado em:

**Banca Examinadora:**

Prof. Dr. Marcelo dos Santos (orientador)

Instituição: Escola de Comunicações e Artes – Universidade de São Paulo (ECA-USP)

Julgamento: \_\_\_\_\_

Prof.(a) Dr. (a): \_\_\_\_\_

Instituição: \_\_\_\_\_

Julgamento: \_\_\_\_\_

À minha mãe.

## **AGRADECIMENTOS**

O desenvolvimento deste trabalho ocorreu em um período com muitos desafios, contratemplos e percalços pessoais. Assim, não seria possível sem a ajuda de várias pessoas, às quais sou imensamente grato.

Primeiramente, agradeço ao Prof. Dr. Marcelo dos Santos, pela orientação, confiança, apoio, incentivo e, principalmente, paciência. Sou muito grato por seus ensinamentos, levarei comigo seu exemplo como pesquisador e cientista.

À Universidade de São Paulo, pelo espaço concedido para o desenvolvimento deste trabalho.

Aos amigos que fiz no programa, os quais dividiram suas ideias, alegrias, frustrações e momentos de descontração.

Ao Reagan, Clinton e Ty, companheiros de madrugada, no desenvolvimento deste trabalho.

À minha esposa, Beatriz. Nada disso seria possível sem sua paciência e insistência. A pesquisa é um trabalho árduo e, por vezes, estafante. Sem seu apoio, eu não teria chegado onde cheguei. Muitas vezes, ela acreditou mais em mim do que eu mesmo acreditava. Por isto, sou eternamente grato.

“- Está bem - disse o computador. - A Resposta para à Grande Questão...

- Sim...!

- Da Vida, o Universo e Tudo Mais... - disse Pensador Profundo.

- Sim...!

- É... disse Pensador Profundo, e fez uma pausa.

- Sim...!

- É...

- Sim...!!!...?

- Quarenta e dois - disse Pensador Profundo, com uma majestade e uma tranquilidade infinitas.”

(ADAMS, 2004, p.132).

## RESUMO

ALBANO, Luís Roberto Momberg. **Modelo para recuperação de informações em ambientes digitais**: uma proposta baseada no contexto de uso. 2023. Doutorado (Ciência da Informação) – Universidade de São Paulo - USP, São Paulo, SP, 2023.

A oferta de informações qualificadas e adaptadas aos diferentes contextos, nos quais são demandadas e utilizadas por diferentes usuários, constitui um desafio para a área de Ciência da Informação. Tal desafio se intensifica quando se trata da disponibilização de informações em ambientes digitais, nos quais o acesso é concedido a grupos heterogêneos de usuários. O mesmo desafio também é amplificado pela veloz, crescente e variada produção de dados e informações, incluindo novas possibilidades de uso por diferentes interessados. Nesse cenário, em função da diversidade de usuários, suas interações, interesses e contextos de utilização das informações disponíveis em tais ambientes, as dificuldades associadas ao projeto e oferta de instrumentos para recuperação de informações também aumentaram. Em parte, isto se deve ao fato de que, tradicionalmente, a seleção, coleta e organização, incluindo o projeto e implementação de instrumentos para recuperação de informações, consideram a figura do potencial usuário. Contudo, no caso de ambientes informacionais digitais disponibilizados ao grande público, nas fases de projeto e implementação, nem sempre é possível identificar o potencial usuário. Assim, neste trabalho, como hipótese, conjecturou-se que a utilização de atributos (informações) do contexto de uso, em tais ambientes, pode aprimorar a recuperação de informações, com maiores precisão, pertinência, relevância e até acessibilidade. Com isto, o objetivo foi propor um modelo que utiliza atributos vinculados ao contexto de uso da informação, com intuito de melhorar as respostas fornecidas por um sistema de recuperação de informações desenvolvido para ambientes informacionais digitais. Trata-se de uma pesquisa de natureza exploratória, baseada no estudo da literatura, proposição de um modelo para recuperação de informações com base no contexto de uso e apresentação de uma possibilidade de aplicação de tal modelo. Analisando a referida possibilidade de aplicação, pôde-se constatar que informações sobre o contexto de uso são essenciais para o desenvolvimento e aprimoramento dos algoritmos computacionais utilizados na recuperação de informações em ambientes digitais.

Palavras-chaves: Sistema de Recuperação de informação. Contexto de uso. Ambientes digitais.

## ABSTRACT

ALBANO, Luís Roberto Momberg. **Modelo para recuperação de informações em ambientes digitais**: uma proposta baseada no contexto de uso. 2023. Doutorado (Ciência da Informação) – Universidade de São Paulo - USP, São Paulo, SP, 2023.

The provision of qualified information tailored to diverse contexts, in which it is requested and utilized by various users, poses a challenge to the field of Information Science. This challenge intensifies when it comes to provide information in digital environments, where access is granted to heterogeneous user groups. The same challenge is also amplified by the rapid, growing, and varied production of data and information, including new possibilities of use by different stakeholders. In this scenario, due to the diversity of users, their interactions, interests, and contexts of information usage in such environments, the difficulties associated with the design and provision of information retrieval tools have also increased. Partly, this is because traditionally, selection, collection, and organization, including the design and implementation of information retrieval tools, consider the potential user. However, in the case of digital informational environments available to the public, during the design and implementation phases, it is not always possible to identify the potential user. Thus, in this work, as a hypothesis, it was conjectured that the use of attributes (information) from the context of use in such environments can enhance information retrieval with greater accuracy, relevance, pertinence, and accessibility. The aim was to propose a model that uses attributes linked to the context of information usage to improve the responses provided by an information retrieval system developed for digital informational environments. This is an exploratory research based on literature study, proposing a model for information retrieval based on the context of use, and presenting a possibility of applying such model. Analyzing the mentioned application possibility, it was observed that information about the context of use is essential for the development and enhancement of computational algorithms used in information retrieval in digital environments.

Keywords: Information retrieval. Context of use. Information systems. Information retrieval systems.

## LISTA DE FIGURAS

Figura 1 – Estrutura hierárquica DIKW conforme Rowley (2007) .....	29
Figura 2 – Estrutura do Comportamento Informacional conforme Wilson (2000) .....	36
Figura 3 - Fluxo interno e os fluxos externos da informação .....	43
Figura 4 - Funções de um sistema de recuperação de informações .....	46
Figura 5 - Modelo de SRI proposto por Ingwersen (1996) .....	48
Figura 6 - Modelo atualizado de SRI de Ingwersen (1999) .....	50
Figura 7 - Fluxo de um sistema de recuperação de informações digital conforme Ferneda (2012) .....	52
Figura 8 - Relação matemática entre os atributos de revocação e precisão em uma recuperação de informações .....	75
Figura 9 - Relação entre os atributos de revocação e precisão em uma recuperação de informações .....	75
Figura 10 - Relação entre a especificidade e a exaustividade da informação recuperada .....	76
Figura 11 - Estrutura da usabilidade de computadores, com indicação de contexto de uso, conforme norma NBR 9241-11 .....	83
Figura 12 - Fluxograma de desenvolvimento do trabalho .....	97
Figura 13 - Bloco da estrutura básica do diagrama IDEF-0 .....	111
Figura 14 – Fluxo informacional em um SRI, conforme Lima e Campos (2022) ....	116
Figura 15 - Fluxo informacional em um SRI, conforme Vickery e Vickery (2004) ...	118
Figura 16 – Diagrama de caso de uso do modelo de SRI proposto no trabalho ....	127
Figura 17 - Processo de operação do modelo proposto representando em diagrama do tipo IDEF-0 .....	130
Figura 18 - Teoria da ação, conforme proposto por Donald Norman (1986) .....	132
Figura 19 - Interface inicial de acesso do usuário ao SRI do modelo proposto .....	136
Figura 20 - Captação de atributos socioculturais a partir de formulário de preenchimento do usuário .....	139
Figura 21 – Modelo de organização de dados em um banco de dados NoSQL do tipo colunar .....	142
Figura 22 - Processo de KDT para captura de atributos contextuais socioculturais e cognitivos do modelo proposto no trabalho .....	146
Figura 23 - Mapa contextual composto por atributos socioculturais e cognitivos capturados em processo de KDT no modelo proposto no trabalho .....	148
Figura 24 - Comparação entre termos vinculadores dos documentos e mapa contextual gerado pela captura de atributos de contexto de uso socioculturais e cognitivos ..	150
Figura 25 - Interface de apresentação da informação recuperada no modelo proposto, apresentando as variedades de formas de apresentação .....	156
Figura 26 – Forma de refinamento dos termos vinculadores dos documentos, a partir do processo de avaliação da informação recuperada .....	160
Figura 27 - Diagrama de telas do fluxo de uso do SRI proposto no trabalho .....	162
Figura 28 - BPMN da operação geral do modelo proposto no trabalho .....	164
Figura 29 – <i>Printscreen</i> da página inicial e campo de pesquisa de referências da BRAPCI .....	174
Figura 30 – Busca em linguagem natural no campo de referências da BRAPCI (considerando a relação CI e o filósofo Habermas) .....	176
Figura 31 - Busca em linguagem natural no campo de referências da BRAPCI (considerando a relação CI e o filósofo Foucault) .....	176

Figura 32 – Resultado de busca de referências para o descritor “Habermas”.....	178
Figura 33 – Métodos de busca avançada incorporadas atualmente ao sistema da BRAPCI.....	180
Figura 34 – Busca de referências dos descritores “Habermas” e “Ciência da Informação” utilizando o operador avançado “AND”.....	181
Figura 35 - Modelo de relacionamento de dados (BRAPCI).....	184
Figura 36 – Proposta de integração do modelo apresentado com o sistema da BRAPCI.....	185
Figura 37 – Proposta de interface de cadastro de usuário na BRAPCI, utilizando o modelo proposto no trabalho.....	191
Figura 38 – Campo de seleção para utilização do processo de busca contextual, conforme modelo apresentado no trabalho.....	192
Figura 39 – Proposta de interface de apresentação de informações recuperadas, conforme modelo apresentado no trabalho.....	195
Figura 40 – Proposta de campo de avaliação da informação recuperada, conforme modelo apresentado no trabalho.....	197
Figura 41 – Proposta de interface de <i>feedback</i> , conforme modelo apresentado no trabalho.....	198
Figura 42 - Fluxo informacional do SRI proposto no trabalho, com base na estrutura de Lima e Campos (2022).....	208

## LISTA DE QUADROS

Quadro 1 - Características gerais de dado, informação e conhecimento .....	30
Quadro 2 - Tipos de SRIs organizados por suporte .....	47
Quadro 3 - Atributos contextuais levantados na elaboração do referencial teórico ..	91
Quadro 4 - Autores e temas utilizados no trabalho .....	99
Quadro 5 - Forma de apresentação da informação recuperada no modelo proposto .....	154
Quadro 6 - Código fonte do programa de teste do modelo proposto no trabalho ...	171
Quadro 7 - Matriz SWOT para avaliação do modelo proposto .....	206

## LISTA DE SIGLAS

ABNT	Associação Brasileira de Normas Técnicas
BRES	Base Brasil/Espanha de artigos de periódicos da área em Ciência da Informação
BD	Banco de Dados
BI	<i>Business Intelligence</i> (Inteligência de Negócio)
BPMN	<i>Business Process Model and Notation</i> (Modelagem e Notação de Processos de Negócios)
BRAPCI	Base de Dados Referencial de Artigos de Periódicos em Ciência da Informação
BSON	<i>Binary JSON</i> (JSON Binário)
CD-ROM	<i>Compact Disc - Read-Only Memory</i> (Disco compacto de somente leitura)
CI	Ciência da Informação
CSS	<i>Cascading Style Sheets</i> (Folhas de Estilo em Cascata)
GPT	<i>Generative Pre-trained Transformer</i> (Transformador Generativo Pré-Treinado)
HCI/IHC	<i>Human-Computer Interaction</i> (Interação Humano-Computador)
HDFS	<i>Hadoop Distributed File System</i> (Sistema Distribuído de Arquivos do Hadoop)
HTML	<i>Hypertext Markup Language</i> (Linguagem de Marcação de Hipertexto)
HTTP	<i>Hypertext Transfer Protocol</i> (Protocolo de Transferência de Hipertexto)
IA	Inteligência Artificial
IDEF-0	<i>Icam DEFinition for Function Modeling</i> (Definição ICAM para Modelagem de Funções)
IOT	<i>Internet of Things</i> (Internet das Coisas)
JSON	<i>JavaScript Object Notation</i> (Notação de Objeto JavaScript)
KDT	<i>Knowledge Discovery in Texts</i> (Descoberta de conhecimento em textos)
ML	<i>Machine Learning</i> (Aprendizado de Máquina)

NBR	Norma Brasileira
NoSQL	<i>Not Only SQL</i> (Não Apenas SQL)
OJS	<i>Open Journal Systems</i> (Sistema de Publicação de Periódicos em Acesso Aberto)
PHP	<i>Hypertext Preprocessor</i> (Pré-processador de Hipertexto)
PUC	Pontifícia Universidade Católica
RI	Recuperação de Informações
SI	Sistema de Informação
SQL	<i>Structured Query Language</i> (Linguagem Estruturada de Consulta)
SRI	Sistema de Recuperação de Informação
SRID	Sistema de Recuperação de Informação Digital
SWOT	<i>Strengths</i> (Forças), <i>Weaknesses</i> (Fraquezas), <i>Opportunities</i> (Oportunidades) e <i>Threats</i> (Ameaças)
TDIC	Tecnologias Digitais de Informação e Comunicação
TI	Tecnologia da Informação
UFRGS	Universidade Federal do Rio Grande do Sul
UNICAMP	Universidade Estadual de Campinas
USP	Universidade de São Paulo
XML	<i>Extensible Markup Language</i> (Linguagem de Marcação Extensível)

## SUMÁRIO

<b>CAPÍTULO 1 – INTRODUÇÃO</b> .....	<b>14</b>
1.1 Contexto e problema de pesquisa .....	16
1.2 Hipótese.....	17
1.3 Objetivos geral e específicos.....	19
1.4 Motivação e justificativa.....	19
1.5 Estrutura e organização do trabalho.....	21
<b>CAPÍTULO 2 – REFERENCIAL TEÓRICO</b> .....	<b>23</b>
2.1 Informação na perspectiva da Ciência da Informação: aspectos gerais ..	25
2.2 Usuários da Informação, sujeitos informacionais: características gerais .	33
2.3 Sistemas de informação e recuperação da informação .....	40
2.4 Aspectos gerais de sistemas de recuperação de informação digital.....	57
2.4.1 Técnicas e métodos de processamento de dados em sistemas de informação digitais.....	61
2.5 Relevância e pertinência da informação e avaliação de sistemas de recuperação da informação .....	67
2.6 A noção de contexto .....	77
2.7 Considerações gerais sobre o estado da arte de sistemas de informação digitais.....	93
<b>CAPÍTULO 3 – PROCEDIMENTOS METODOLÓGICOS</b> .....	<b>95</b>
<b>CAPÍTULO 4 – MODELO PARA RECUPERAÇÃO DE INFORMAÇÕES</b> .....	<b>113</b>
4.1 Considerações gerais da proposição do modelo.....	114
4.2 Modelo proposto para recuperação da informação em ambientes informacionais digitais .....	120
4.3 Detalhamento do modelo proposto.....	128
4.3 Exemplo de utilização do modelo proposto .....	165
4.3.1 Experimento de operações técnicas básicas do modelo .....	169
4.3.2 Características do site da BRAPCI .....	173
4.3.3 Integração do modelo de SRI proposto ao site da BRAPCI .....	182
4.3.2 Operação do modelo de SRI proposto na BRAPCI .....	190
4.3.4 Considerações sobre a integração do modelo ao site da BRAPCI .....	199
<b>CAPÍTULO 5 – DISCUSSÃO</b> .....	<b>203</b>
5.1 Abordagem do modelo em relação ao problema de pesquisa e à hipótese .....	204

5.2 Discussão .....	207
<b>CAPÍTULO 6 – CONCLUSÕES E CONSIDERAÇÕES FINAIS.....</b>	<b>212</b>
6.1 Conclusões .....	215
6.2 Limitações da pesquisa .....	220
6.3 Trabalhos futuros .....	222
<b>Referências .....</b>	<b>225</b>
<b>APÊNDICES .....</b>	<b>234</b>
Apêndice 1 .....	235
Apêndice 2 .....	236
Apêndice 3 .....	237
Apêndice 4 .....	238
Apêndice 5 .....	239

## **CAPÍTULO 1 – INTRODUÇÃO**

Contemporaneamente, um dos desafios da Ciência da Informação (CI) está associado à recuperação e oferta de informações com precisão, relevância, pertinência e acessibilidade, para quem, quando, onde e da forma que se necessita. Em outras palavras, tal desafio se relaciona com a oferta de informações ditas qualificadas, sobretudo em uma época em que se produz e se fazem necessárias ferramentas próprias para a seleção, filtragem e apresentação dessas informações, de forma tida como adequada a cada um dos diferentes contextos em que são demandadas.

Nesse cenário e de uma forma bem geral, os ambientes informacionais digitais (ou sistemas de informação) contemporâneos tem potencializado o referido desafio. O que, dessa forma, impacta no projeto e implementação de sistemas de recuperação de informações. Tais sistemas, em geral, materializam estratégias para obter (recuperar) informações armazenadas em diferentes repositórios ou bases de dados. Pois, de um modo bastante sucinto, um sistema de recuperação de informações (SRI) pode ser visto como um instrumento projetado para realizar um conjunto de

[...] operações interligadas para identificar, dentre um grande conjunto de informações (uma base de dados, por exemplo), aquelas que são de fato úteis, ou seja, que estão de acordo com a demanda expressa pelo usuário (ARAÚJO JÚNIOR, 2005, p. 77).

Com isto, compreende-se que os SRIs são projetados para oferecer respostas a perguntas de usuários. Por um lado, contudo, tradicionalmente, os mesmos sistemas são projetados levando-se em consideração a figura de seu potencial usuário. Por outro, os ambientes informacionais digitais modernos, usualmente, têm sido projetados sem ter como referência a(s) figura(s) do(s) potencial(is) usuário(s). Logo, observa-se que oferecer respostas adequadas às demandas dos diversos usuários, de forma qualificada, constitui um desafio. E tal desafio, visto como uma barreira para acesso e uso de informações, tende a impactar o desenvolvimento de tarefas do cotidiano de pessoas e organizações.

Usualmente, no caso do projeto de sistemas de informação (SIs) e SRIs ditos clássicos<sup>1</sup>, a partir do perfil do potencial usuário, se faz a identificação e a seleção das

---

<sup>1</sup> SRIs clássicos, neste trabalho, são entendidos como [...] um conjunto de componentes que armazenam, processam e distribuem informação (PONJUÁN-DANTE, 2006, p. 22, tradução nossa). [...] un conjunto de componentes que almacenan, procesan y distribuyen información (PONJUÁN-DANTE, 2006, p. 22)

informações que serão coletadas e armazenadas em seus repositórios. Nesse caso, previamente ao armazenamento, a descrição e especificação de metadados são planejadas, de modo que tais metadados representem pontos de acesso para a recuperação de tais informações. Isto, normalmente, tem como base o perfil do potencial usuário e possíveis cenários de uso (ou formas de interação do potencial usuário com as informações demandadas). Algo que, tradicionalmente, é idealizado durante o estudo de requisitos do sistema na fase de projeto. Nota-se, dessa forma, que os processos de seleção, coleta e organização de informações consideram um determinado potencial de uso.

De uma forma geral, pode-se notar que a informação considerada útil é a que “responde” à demanda do usuário. No entanto, esta demanda é circunstancial e surge em um dado tempo, espaço e condições que norteiam o uso. Portanto, percebe-se que, recuperar a informação útil e necessária para um determinado usuário, o qual está inserido em um dado contexto, pressupõe a identificação e compreensão de uma série de variáveis e condicionantes que compõem o contexto do usuário (ou contexto de uso da informação).

### **1.1 Contexto e problema de pesquisa**

Conforme já mencionado, contemporaneamente, observa-se a necessidade de desenvolver métodos e instrumentos que possam potencializar a recuperação de informações disponíveis nos variados repositórios. Tais informações foram produzidas, coletadas, organizadas e até processadas nas mais variadas condições e com distintas intenções, utilizando-se de tecnologias digitais de informação e comunicação (TDICs). Adicionalmente, percebe-se que as referidas informações têm o potencial de oferecer suporte ao desenvolvimento de atividades do cotidiano de pessoas e instituições, tendo em vista a tomada de decisões. Nesse sentido, com base nas características dos diferentes cenários, o processo de tomada de decisões tende se tornar mais complexo, exigindo métodos e instrumentos mais sofisticados e capazes de subsidiar tal processo.

Cabe lembrar que uma parte significativa de dados e informações com potencial de subsidiar a tomada de decisões é gerada por equipamentos digitais (incluindo os dispositivos móveis) conectados à internet, os quais registram atividades

e interações de seus usuários (MARQUESONE, 2016). Destaca-se que esses dados e informações são registrados e compartilhados em diferentes formatos (por exemplo, textos, imagens, vídeos e sons), dentre os quais incluem, por exemplo, localização, tempo de uso dos dispositivos e, em alguns casos em que não se adota as devidas políticas de privacidade, identificação do usuário.

Em geral, praticamente tudo o que é produzido em tais ambientes informacionais é coletado e armazenado com a intenção de alguma utilização futura. No entanto, esta coleta é marcada pela ausência de diretrizes (ou políticas) de seleção, as quais poderiam nortear a implementação de estratégias para oferta e utilização do que for coletado. Assim, conforme já mencionado e como estratégia de auxiliar no aprimoramento da recuperação de informações, observa-se o desenvolvimento e implementação de algoritmos computacionais que utilizam técnicas de aprendizagem de máquina (ou *machine learning* – ML), em função do volume de informações coletadas e armazenadas.

Reiterando o que já fora mencionado, o projeto e implementação de SIs levam em conta o potencial usuário, e este usuário é base para a especificação do que será coletado, armazenado e como será apresentado. Por outro lado, SIs digitais contemporâneos, às vezes, são projetados e implementados sem ter a referência do potencial usuário, o que dificulta o projeto e implementação de instrumentos para recuperação da informação. Portanto, nesse cenário, visualiza-se uma certa inflexão no que diz respeito a abordagens utilizadas para implementar e manter SRIs na contemporaneidade. Isto ocorre devido à diversidade de dados e informações coletados, bem como, às variadas possibilidades de uso dos mesmos.

Assim, compreendendo que, em geral, dados e informações são produzidos em um contexto e utilizados em outro(s), a questão norteadora utilizada para o desenvolvimento deste trabalho se resume a *como promover a recuperação de informações em ambientes digitais, tendo em vista que tais ambientes são diferentes dos “tradicionais” sistemas de informação?*

## **1.2 Hipótese**

Dentre outros, observa-se que relevância e pertinência são dois atributos vinculados à qualidade das respostas oferecidas pelos SRIs. Tais atributos estão diretamente relacionados com a adequação da informação obtida e demandada por

um dado usuário em um determinado contexto. Assim, conjecturou-se que o contexto do usuário (ou contexto de uso da informação) pode oferecer subsídios importantes para o aprimoramento da relevância e pertinência das respostas fornecidas por SRIs em ambientes digitais.

Com isto, parte-se do pressuposto que SRIs “sensíveis” ao contexto (*context-aware*) do usuário demandante da informação constituem instrumentos de grande valia, principalmente, em ambientes informacionais digitais. Em termos de contexto de uso, de uma forma geral, considera-se a “combinação de usuários, objetivos e tarefas, recursos e ambientes” (ABNT, 2021, p. 4). Em geral, o contexto inclui elementos como idioma, localização, elementos culturais e outros aspectos, conforme Abowd e Dey (1999) identificaram.

Em termos de **hipótese**, trabalhou-se com a utilização de atributos (ou informações) relacionados com o contexto de uso<sup>2</sup> para refinar os procedimentos de recuperação de informações, uma vez que tal contexto pode oferecer subsídios importantes para “moldar” a resposta que os SRIs oferecerão. Nesse caso, vislumbrou-se que tais subsídios estão vinculados a três propriedades gerais da informação: (1) forma; (2) conteúdo; e (3) acessibilidade (tanto física, quanto cognitiva).

Conhecendo-se, assim, o contexto no qual se encontra o usuário de uma dada informação, poder-se-á apresentar o que for recuperado por um SRI com maiores precisão, relevância e pertinência. Imagina-se que tal apresentação terá como base atributos como: (a) quem é o usuário demandante da informação a ser recuperada; (b) os objetivos e tarefas que pretende realizar com essa informação; (c) recursos (por exemplo, equipamento disponível para apresentar a informação – *tablet*, *smartphone* ou computador pessoal); e (d) a maneira como o conteúdo recuperado será apresentado (forma), considerando particularidades do usuário (promoção das acessibilidades física e cognitiva).

---

<sup>2</sup> Por contexto de uso, neste trabalho, entende-se como o conjunto de “[...] usuários, tarefas, equipamento (*hardware*, *software* e *materiais*), e o ambiente físico e social no qual um produto é usado” (ABNT, 2021, p. 3).

### 1.3 Objetivos geral e específicos

Com base no problema e hipótese, o **objetivo geral** proposto para o desenvolvimento deste trabalho consiste em, frente à compreensão dos desafios presentes nos processos de recuperação de informações, propor um modelo que utiliza elementos do contexto de uso, com vistas ao aprimoramento das respostas fornecidas por um SRI utilizado em ambientes informacionais digitais.

Como **objetivos específicos**, foram estabelecidos:

- a) identificar e compreender, na perspectiva da CI, as características dos ambientes informacionais digitais contemporâneos;
- b) identificar variáveis inerentes ao contexto de uso, as quais poderão auxiliar no aprimoramento de respostas fornecidas por SRIs.

### 1.4 Motivação e justificativa

A informação, como se sabe, representa um insumo de grande valia para o desenvolvimento de diversas atividades do cotidiano, tanto de pessoas, quanto de organizações. No entanto, apesar da ampla utilização das TDICs para auxiliar na produção, coleta e armazenamento de informações, ainda existem barreiras, especialmente no que diz respeito à oferta de informações adequadas a cada usuário. Com isto, a **motivação** para o desenvolvimento deste trabalho se assentou na necessidade de revisitar e repensar abordagens utilizadas na recuperação de informações, considerando o amplo uso das TDICs e SIs digitais.

Atualmente, observa-se o desenvolvimento de diversas tecnologias que implementam abordagens variadas para minimizar os problemas da recuperação de informações. Por exemplo, a tecnologia *Generative Pre-trained Transformer* (GPT) oferece um modelo e infraestrutura para sistemas capazes de realizar a geração de textos, imagens e outros registros em resposta a uma dada demanda. Com isto, contemporaneamente, ferramentas baseadas em GPT, como o ChatGPT, representam avanços no que diz respeito à utilização de informações contextuais para responder de forma “mais adequada” às demandas de usuários. Por outro lado, tais tecnologias também apresentam desafios para serem superados. Dentre os quais, estão possíveis vieses, falta de transparência, manipulação de informações e dependência de grandes conjuntos de dados.

Diversos trabalhos e pesquisas desenvolvidos na perspectiva da CI e Ciência da Computação têm abordado os problemas da recuperação de informações, privilegiando o estudo de métodos e procedimentos. No entanto, apesar oferecerem contribuições relevantes, parte dos trabalhos não colocam no centro das investigações a figura do potencial usuário e, conseqüentemente, o seu contexto. Por outro lado, trabalhos como os de Dey (2000), Talja, Heidi e Tarja (1999), Foresti, Varvakis e Godoy Viera (2016), Henrique, Nassif e Venâncio (2007) e Johnson (2003) consideraram o contexto do usuário como uma variável a ser ponderada no que diz respeito ao acesso e uso de informações em diferentes sistemas. Contudo, no caso dos ambientes informacionais digitais contemporâneos, não é possível vislumbrar, durante a fase de projeto, os potenciais usuários e possíveis interações que estes realizarão com o sistema. Portanto, isto exige mudanças em termos de estratégias para projeto de implementação de SIs e, sobretudo, SRIs, em um momento em que as diversas TDICs estão presentes e influenciam diretamente o cotidiano de pessoas e instituições.

Em termos de **justificativa**, destaca-se a urgência por compreender e ressignificar “[...] os estudos das necessidades informacionais, do estudo do fluxo e uso da informação” (TARAPANOFF, 2006, p. 19), frente aos avanços tecnológicos subsidiados pelas TDICs, sobretudo, em termos de SIs digitais. Embora, particularmente com tais sistemas, também é almejado agregar valor por meio da oferta de informações para indivíduos ou grupos (DE MAURO; GRECO; GRIMALDI, 2016), nem sempre isso acontece como o esperado. Pois, imagina-se que tais sistemas consigam estruturar e apresentar informações de maneira adequada para atender às demandas dos usuários. Contudo, na prática, os referidos sistemas falham em tal tarefa, o que evidencia barreiras de dimensões consideráveis na tentativa de assegurar acesso e uso de informações em diferentes contextos.

Isso traz certo grau de preocupação, considerando que os SIs podem fornecer respostas bastante imprecisas e que podem impactar a tomada de decisões. A falta de estudos e investigações com vistas a mitigar o referido problema contribui para potencializar as barreiras de acesso e uso de informações disponíveis em ambientes digitais. Por exemplo, além dos problemas de recuperação de informações, SRIs limitados e mal planejados, direta ou indiretamente, podem contribuir com o agravamento de problemas sociais, por exemplo, a discriminação de pessoas e a disseminação de notícias falsas, entre outros (O’NEIL, 2021; ZUBOFF, 2021).

Também, do ponto de vista teórico-metodológico, a compreensão das dinâmicas de SIs planejados sem referência ao potencial usuário representa uma mudança de paradigma na concepção destes sistemas. Especialmente, em relação a SIs destinados ao trabalho com grandes volumes, variedades e velocidade em termos de produção informacional. Portanto, identifica-se a necessidade de compreender melhor as dinâmicas presentes nas interações usuário–informação, de modo que as soluções disponibilizadas em termos de produtos e serviços de informação possam ser de fato adequadas a cada grupo de usuários, considerando-se as características específicas de seu contexto.

Em síntese, compreende-se que, em termos de recuperação de informações, se faz necessário examinar mais de perto os aspectos relacionados aos diferentes contextos de uso dos que trabalham e dependem de informações em ambientes informacionais digitais. Em especial, os SIs digitais e outros sistemas disponibilizados com a utilização das TDICs evidenciam uma série de desafios para serem superados, particularmente, no que diz respeito à concepção do projeto e implementação de métodos e instrumentos de recuperação de informações, e a CI aponta caminhos, pois como observado por Tarapanoff (2006), a área trata do “[...] estudo das propriedades gerais da informação (natureza, gênese e efeitos)” (TARAPANOFF, 2006, p. 18-19).

### **1.5 Estrutura e organização do trabalho**

O presente trabalho está organizado em seis capítulos. Sendo este a Introdução, no qual foram apresentados o contexto e problema de pesquisa, hipótese, objetivos geral e específicos, incluindo a motivação e justificativa para desenvolvimento do estudo.

No segundo capítulo, Referencial Teórico, foram apresentados os aportes teóricos e metodológicos que sustentaram o desenvolvimento de tal estudo. O referido capítulo foi dividido em sete seções, nas quais foram apresentados diversos aspectos relacionados à CI, usuários da informação, sistemas de informação, recuperação de informações e sistemas de informação digital. O principal intuito foi o de compreender a dinâmica de operação em tais ambientes informacionais digitais e explorar como o contexto de uso e os perfis dos usuários podem ser utilizados para aprimorar tais

sistemas. Na última seção, apresenta-se uma reflexão sobre o potencial de utilização do contexto de uso e perfil de usuário para aprimoramento dos SRIs.

Seguindo, no terceiro capítulo, Procedimentos Metodológicos, foram apresentadas as estratégias adotadas para a condução do estudo. Em resumo, o trabalho se desenvolveu em três etapas principais: (1) estudo de literatura e preparação do referencial teórico; (2) desenvolvimento de um modelo conceitual para recuperação da informação em ambientes informacionais digitais; e (3) exposição de uma possível aplicação do modelo proposto.

Na sequência, no quarto capítulo, Proposta de um Modelo para Recuperação de Informações, foi apresentado o modelo desenvolvido, bem como uma possibilidade de utilização deste.

No quinto capítulo, os resultados obtidos foram analisados à luz do referencial teórico preparado e utilizado neste trabalho.

Por fim, no sexto capítulo, Conclusão e Considerações Finais, estão apresentadas as conclusões, contribuições do estudo, limitações e indicações de trabalhos futuros.

## **CAPÍTULO 2 – REFERENCIAL TEÓRICO**

Neste capítulo, estão apresentados os fundamentos teóricos e metodológicos utilizados no desenvolvimento deste trabalho. O capítulo foi estruturado em sete seções, buscando apresentar o objeto de estudo na perspectiva da CI, identificar os métodos e procedimentos empregados no estudo e fornecer subsídios para sustentar as reflexões acerca dos resultados obtidos.

Na seção 2.1, foram apresentados os aspectos gerais da informação na perspectiva da CI. Entende-se que o termo informação comporta certa ambiguidade. Sendo assim, foram apresentadas a conceituação e caracterização da informação, tendo em vista que o conceito de informação para a CI traz, implicitamente em sua definição, a figura do potencial usuário.

Dessa forma, na seção 2.2, estão apresentadas as considerações gerais sobre usuários da informação, bem como estão explicitados os conceitos de necessidade informacional, comportamento informacional e sujeito informacional. Tais conceitos evidenciam a forma como determinado usuário (ou grupo de usuários) atua quando necessita de informação.

Seguindo, na seção 2.3, foi apresentada uma síntese relativa aos conceitos de SIs e SRIs, apresentando modelos que consideram fluxos informacionais, contemplando seleção, coleta, organização, armazenamento e disponibilização da informação.

Na seção 2.4, foram abordados os sistemas informacionais digitais, com ênfase em seu funcionamento, desafios e as complexidades envolvidas na obtenção de informações relevantes e pertinentes para potenciais usuários.

Já na seção 2.5, foram apresentados os conceitos de relevância e pertinência da informação, ressaltando suas interconexões com o que é recuperado e considerado valioso por um potencial usuário. A partir de tais noções, foram apresentados métodos de avaliação para sistemas de recuperação de informações, tais como precisão e revocação.

O Referencial Teórico continuou com a apresentação de definições de contexto de uso da informação, explorando diferentes abordagens de autores da CI e de áreas correlatas, como Tecnologia da Informação (TI), Psicologia, Sociologia, entre outras.

Por fim, na seção 2.7, foi apresentada uma reflexão sobre o potencial de uso do contexto para aprimorar as respostas fornecidas pelos SRIs.

## 2.1 Informação na perspectiva da Ciência da Informação: aspectos gerais

Conforme observado por Buckland (1991), o termo "informação" pode ter diferentes conotações, dependendo do campo de conhecimento e do contexto em que é utilizado.

Nesse sentido, antes de apresentar o conceito de informação utilizado neste trabalho, observou-se a importância de identificar alguns dos limites de atuação da CI. Borko (1968) apresentou uma definição de CI, considerando a relação desta com a documentação e a computação. Em tal definição, CI é o campo que

[...] investiga as propriedades e comportamento da informação, as forças que regem os fluxos informacionais e os meios de processamento da informação para obter acessibilidade e usabilidade ótimas. Se preocupa com um conjunto de conhecimentos relacionados à origem, coleta, organização, armazenamento, recuperação, interpretação, transmissão, transformação e utilização da informação. Isso inclui a investigação das representações da informação em sistemas naturais e artificiais, o uso de códigos ou transmissão eficiente de mensagens, e o estudo de dispositivos e técnicas de processamento de informação, como computadores e seus sistemas de programação. É uma ciência interdisciplinar derivada e relacionada a campos como Matemática, Lógica, Linguística, Psicologia, Tecnologia Da Computação, Pesquisa Operacional, Artes Gráficas, Comunicações, Ciência Da Informação, Gestão e outros campos similares (BORKO, 1968, p. 1, tradução nossa)<sup>3</sup>.

Nota-se que as pesquisas realizadas no período em que Borko (1968) introduziu seu conceito de informação tinham como foco principal o objeto da atividade de produção, organização e busca de informação. Naquele período, o enfoque das pesquisas em sistemas de informação estava primordialmente voltado para a operação técnica destes sistemas. A visão predominante era a de que o conhecimento seria um "ente" objetivo e especializado, muitas vezes desvinculado do usuário que o

---

<sup>3</sup> "[...] investigates the properties and behavior of information, the forces governing the flow of information, and the means of processing information for optimum accessibility and usability. It is concerned with that body of knowledge relating to the origination, collection, organization, storage, retrieval, interpretation, transmission, transformation, and utilization of information. This includes the investigation of information representations in both natural and artificial systems, the use of codes or efficient message transmission, and the study of information processing devices and techniques such as computers and their programming systems. It is an interdisciplinary science derived from and related to such fields as mathematics, logic, linguistics, psychology, computer technology, operations research, the graphic arts, communications, library science, management, and other similar fields" (BORKO, 1968, p. 1).

acessava. Profissionais de áreas como engenharia, matemática e física adotavam uma perspectiva tecnicista, ajustada às suas demandas específicas.

Essa orientação pode ser vista como "tecno-centrista", pois o mecanismo de busca e recuperação de informação era tratado como um processo determinista e estático. Aspectos relacionados ao uso da informação, como fatores psicológicos e contexto do usuário, muitas vezes eram negligenciados, dando prioridade à operação do sistema em si. Em contraponto, Neveling e Wersig (1975) apresentaram outra definição de CI, a qual inclui a figura do potencial usuário como elemento de estudo. Neste caso, explicitaram que a CI

[...] preocupa-se com a organização dos processos de comunicação direcionados à informação para uma clientela específica. Esta é uma ciência que se assemelha um pouco à comunicação de massa, com o objetivo de atender às necessidades de informação socialmente e individualmente justificáveis de um grande público (NEVELING; WERSIG, 1975, p. 34, tradução nossa).<sup>4</sup>

Ao introduzir a noção de "clientela específica", Neveling e Wersig (1975) sugeriram que os estudos em CI deveriam considerar as necessidades de grupos específicos que utilizam a informação de maneira particular. Ao destacar o papel do usuário da informação, Neveling e Wersig (1975) expuseram que o usuário é um elemento importante para a análise dos processos informacionais e dos fluxos de informação. A definição apresentada por Neveling e Wersig (1975) se assemelha àquela apresentada por Tarapanoff (2006) e é complementar à definição apresentada por Borko (1968). Tarapanoff (2006) apontou que a CI

[...] é uma ciência de caráter eminentemente interdisciplinar, que tem por objeto o estudo das propriedades gerais da informação (natureza, gênese e efeitos). Assim, em pesquisas que abordam o tema da informação, a Ciência da Informação contribui principalmente com estudos das necessidades informacionais, do estudo do fluxo e uso da informação. (TARAPANOFF, 2006. p. 19).

O fluxo informacional, citado por Tarapanoff (2006), diz respeito à forma como as informações são coletadas, processadas e distribuídas em um sistema, incluindo as pessoas envolvidas, os recursos necessários e os mecanismos utilizados para

---

<sup>4</sup> "concerned with the organization of communication processes aimed at information for a specific clientele. This is a science somewhat similar to mass communication aimed at the fulfilment of the socially and individually justifiable information needs of a large public" (NEVELING; WERSIG, 1975, p. 34).

transmitir a informação. Isso pressupõe, portanto, a existência de canais de comunicação específicos para a informação (NEVELING; WERSIG, 1975). De forma complementar, Smit e Barreto (2002) definiram CI como

[...] campo que se ocupa e se preocupa com os princípios e práticas da criação, organização e distribuição da informação, bem como com o estudo dos fluxos da informação desde sua criação até a sua utilização, e sua transmissão ao receptor em uma variedade de formas, por meio de uma variedade de canais. (SMIT; BARRETO, 2002, p. 17).

Smit e Barreto (2002) não se limitaram ao estudo dos fluxos de informação. Também, se preocuparam com os métodos de organização e distribuição desta, bem como, com os canais de transmissão para o receptor. A definição apresentada por eles se assemelha àquela proposta por Le Coadic (2004, p. 17), o qual entende a CI como um campo que se dedica ao estudo das “propriedades da informação, como a sua construção, comunicação e uso”, levando em conta tanto os fluxos de comunicação, quanto o uso da informação por parte do usuário. Ademais, Smit (2012) apontou que pesquisas em CI também levam em conta questões referentes aos contextos em que a informação é produzida e utilizada, reconhecendo que tais contextos são influenciados pelas particularidades sociais dos indivíduos.

Nesse sentido, Smit (2012) indicou que a CI,

[...] se concebe, de fato, como uma ciência social, muito influenciada pelas tecnologias da informação e da comunicação (as TICs) e inserida nos propósitos da sociedade da informação. Obviamente, essa visão epistêmica do uso social da informação abre espaço para novas discussões acerca do papel e do poder da informação e dos sistemas de informação que, ao organizarem a informação, impõem aos usuários determinada visão de mundo. A compreensão dos contextos de produção e uso da informação (contextos social, organizacional ou profissional) constitui uma condição necessária para o trabalho com a informação (SMIT, 2012, p. 90).

Tal abordagem considerou, não apenas as mensagens presentes nos documentos que são utilizadas por algum indivíduo, mas, também, o conhecimento dos contextos, motivações e intenções, sobre os quais a informação foi produzida e pode ser buscada e interpretada. No contexto do desenvolvimento deste trabalho, compreendeu-se que a abordagem descrita por Smit (2012) apresenta maior aderência aos objetivos e problema de pesquisa propostos, bem como, ao objeto de estudo.

Portanto, entende-se que a investigação de sistemas em contextos digitais é de suma importância para a CI. Isto se deve não apenas ao fato de que tais sistemas atuam na gestão e organização da informação, mas, também, porque exercem um impacto social significativo. Além disso, os mesmos sistemas servem como instrumentos informacionais, ou seja, como mecanismos especializados que facilitam o acesso, a disseminação e a utilização eficaz da informação para diversos públicos e contextos.

Observa-se que, contemporaneamente, o desenvolvimento de tais sistemas têm seu projeto e desenvolvimento fortemente alicerçados em algoritmos computacionais. Em termos de definição, um algoritmo computacional é visto como um conjunto finito de instruções bem definidas e ordenadas, que descrevem uma sequência de ações que devem ser executadas por um computador para resolver um problema, ou realizar uma tarefa específica (CORMEN; LEISERSON; RIVEST; STEIN, 2009). Nos sistemas de informação digitais, algoritmos computacionais são implementados com o propósito de organizar e apresentar dados e informações aos usuários.

Neste caso, os algoritmos computacionais são programados para organizar e apresentar dados e informações aos usuários, com base em métodos pré-estabelecidos para acesso, busca e recuperação de informação (ROWLEY, 2007). Os mesmos algoritmos atuam como mecanismos que viabilizam diversas funções nos sistemas, desde a busca até a apresentação de informações ao usuário.

Rowley (2007), considerando uma abordagem voltada TI, esclarece que a informação é essencialmente dados organizados que adquirem sentido e valor para um usuário específico. Esta concepção coincide com a apresentada por Davenport (2014) no âmbito da Gestão de Negócios, o qual sugere que sistemas de informação digitais têm o potencial de criar valor para os usuários ao gerar conhecimento a partir das informações processadas.

A estrutura hierárquica DIKW (*data, information, knowledge, wisdom*), apresentada por Rowley (2007), serve para ilustrar a relação entre dados e informação no contexto da TI. Segundo Rowley (2007), dados são representações simbólicas que se tornam informação quando interpretadas e organizadas de forma a responder a perguntas específicas ou permitir ações informadas. A figura 1 apresenta esta estrutura hierárquica.

**Figura 1 – Estrutura hierárquica DIKW<sup>5</sup> conforme Rowley (2007)**



Fonte: Rowley (2007, p. 164, tradução nossa).

A distinção entre dados e informação é funcional, não estrutural, como apontado por Rowley (2007). Ambos podem ser armazenados ou apresentados de maneiras semelhantes, mas seu valor e uso diferem. SIs desempenham um papel importante na coleta, armazenamento, processamento e apresentação desses elementos aos usuários, por meio de interfaces projetadas para mediar a interação entre o usuário e o sistema.

Por outro lado, conhecimento e sabedoria são entidades tácitas e não são acessíveis de forma explícita por sistemas de informação, conforme delineado por Davenport e Prusak (2003). Estes elementos são gerados na mente do usuário e incluem reflexão, síntese e contexto. A relação entre dados, informação, conhecimento e sabedoria é resumida no Quadro 1, conforme organizado por Davenport e Prusak (2003, p. 18).

---

<sup>5</sup> *Hierarquia de Data, Information, Knowledge e Wisdom, tradução nossa de Rowley (2007, p. 164).*

**Quadro 1 - Características gerais de dado, informação e conhecimento**

<b>Dados</b>	<b>Informação</b>	<b>Conhecimento</b>
<p>Simples observações sobre o estado do mundo</p> <ul style="list-style-type: none"> <li>• Facilmente estruturado;</li> <li>• Facilmente obtido por Máquinas;</li> <li>• Frequentemente quantificado;</li> <li>• Facilmente transferível.</li> </ul>	<p>Dados dotados de relevância e propósito</p> <ul style="list-style-type: none"> <li>• Requer unidade de análise;</li> <li>• Exige consenso em relação ao significado;</li> <li>• Exige, necessariamente, a mediação humana.</li> </ul>	<p>Informação valiosa da mente humana inclui reflexão, síntese, contexto</p> <ul style="list-style-type: none"> <li>• De difícil estruturação;</li> <li>• De difícil captura em máquinas;</li> <li>• Frequentemente tácito;</li> <li>• De difícil transferência.</li> </ul>

Fonte: Davenport e Prusak (2003, p. 18).

A referida diferenciação de dados, informação e conhecimento, explicitadas por Davenport e Prusak (2003), tem como origem o campo da Administração e Gestão, porém, guardam relação com o que foi descrito por Buckland (1991), sobre as três dimensões do termo "informação", no que se refere à sua condição tangível e intangível, como elementos manuseável ou elemento mental: (1) como coisa, que pode ser representada por objetos ou artefatos que transmitem informações, como, por exemplo, um documento, em que a informação é explícita; (2) como processo, referindo-se à comunicação de um determinado fato; e (3) como conhecimento, que tem o potencial de reduzir a incerteza de uma pessoa.

Em uma análise interdisciplinar, é essencial reconhecer que o conceito de "informação" é interpretado de maneiras distintas em diferentes campos do saber. Estas definições, como visto, são observadas nas áreas como TI e Administração, onde a informação é frequentemente tratada como um conjunto de dados processados, úteis para a eficiência operacional e tomada de decisões estratégicas. Para a CI, as preocupações com a informação tendem a ser mais abrangentes, abordando aspectos como o ciclo de vida da informação, a comunicação humana e a organização do conhecimento. Nesse sentido, Smit e Barreto (2002, p. 21-22) definiram informação como

[...] estruturas simbolicamente significantes, codificadas de forma socialmente decodificável e registradas (para garantir permanência no tempo e portabilidade no espaço) e que apresentam a competência de gerar conhecimento para o indivíduo e para o seu meio. Estas estruturas significantes são estocadas em função de um uso futuro, causando a institucionalização da informação. (SMIT; BARRETO, 2002, p. 21-22).

Na mesma linha, Le Coadic (2004, p. 4) destacou que a “informação comporta um elemento de sentido”, possuindo “um significado transmitido a um ser consciente por meio de uma mensagem inscrita num suporte espacial-temporal: impresso, sinal elétrico, onda sonora, etc.”. Nota-se que ambas definições reforçam a necessidade de um registro acessível e decodificável, para um grupo específico, a partir de um sistema de signos (LE COADIC, 2004), como visto também na definição apresentada por Rowley (2007), a qual sugere necessidade de um sistema de suporte para os dados e informações.

Além disso, ambas definições assumem a presença de um usuário da informação, ao descreverem que tais estruturas devem ser organizadas de maneira compreensível de forma simbólica. Assim, o usuário pode empregar a informação para atender a diversas necessidades e demandas específicas. Tais necessidades por informações podem ser influenciadas tanto por fatores internos quanto externos (FIGUEIREDO, 1994). Internamente, as mesmas demandas são moldadas por conhecimentos prévios, desejo por conhecimento, valores, crenças e interesses individuais (FIGUEIREDO, 1994). Externamente, as demandas são impulsionadas por situações cotidianas, problemas complexos, bem como mudanças sociais, econômicas e tecnológicas. Tais fatores internos e externos desempenham um papel significativo na determinação das necessidades de informação das pessoas, orientando suas buscas e influenciando o processo de obtenção de conhecimento (FIGUEIREDO, 1994).

Dessa forma, vislumbra-se que a coleta de informações em um SI é orientada pelas potenciais necessidades do usuário, visando contribuir eficazmente para atender tais necessidades. Logo, o principal objetivo de tal abordagem é oferecer informações que possam efetivamente satisfazer a necessidade informacional do potencial usuário. Sobre isso, Smit e Barreto (2002, p. 15) observaram que, para a informação

[...] intervir na vida social, gerando conhecimento que promova o desenvolvimento, [...] necessita ser transmitida e aceita como tal. Os espaços sociais não são homogêneos como o processamento técnico dos estoques de informação. A realidade, em que se pretende que a informação atue, é multifacetada e formada por micronúcleos sociais com divergências profundas. Os habitantes destas comunidades sociais diferenciam-se por condições como: grau de instrução, nível de renda, religião, raça, competência para a codificação e decodificação do código linguístico comum, acesso e interpretação dos códigos formais de conduta moral e ética, acesso à informação que é disponibilizada, confiança no canal de transferência, entre

outros. A uma produção de informação globalizante e globalizada contrapõe-se, inevitavelmente, a recepção local: esta diferenciação, certamente, condiciona a distribuição da informação, o seu uso e a sua assimilação para gerar conhecimento. (SMIT, BARRETO, 2002, p. 15).

Nesse cenário, torna-se imperativo o registro formal das informações. Segundo Smit (2012), informações que são devidamente registradas em algum meio, tecnologia ou código tornam-se mais "estáveis" e "acessíveis" para compartilhamento. A ausência desse registro formal limita a disseminação da informação, independentemente de sua relevância, devido a restrições de tempo e espaço. Portanto, o registro adequado da informação não é apenas uma etapa técnica, mas uma condição necessária para que ela possa ser eficazmente distribuída, utilizada e assimilada, contribuindo para a geração de conhecimento em contextos sociais diversos.

Ao longo do tempo, observou-se a necessidade de criar formas de armazenar e disponibilizar informações registradas para diferentes grupos sociais. Isso levou ao surgimento de instituições como bibliotecas, museus e arquivos, e posteriormente, centros e sistemas de documentação ou informação (SMIT, 2012). Essas instituições são responsáveis por preservar e compartilhar informações e cultura, garantindo que estejam acessíveis e disponíveis por um longo tempo.

Nesses espaços institucionais, políticas abrangentes de gerenciamento da informação são implementadas, abarcando desde a identificação e classificação rigorosa de recursos informacionais até o desenvolvimento de sistemas avançados de gerenciamento de conteúdo. Essas políticas, que estabelecem diretrizes claras para a organização estruturada das informações, são complementadas por investimentos em sistemas que asseguram a integridade, confidencialidade e disponibilidade dos dados. Tais sistemas também facilitam o controle de versões e o compartilhamento seguro de informações dentro da organização. A eficácia dessas políticas e sistemas é maximizada com base na capacitação contínua dos funcionários em ferramentas e tecnologias cruciais para o gerenciamento adequado da informação. Adicionalmente, a decisão de preservar informações específicas, avaliadas como potencialmente úteis, é influenciada pelas condições culturais vigentes no momento da decisão. Conforme apontado por Smit (2012), o "poder informativo" de um fato, evento ou registro é contextual, sendo determinado pelo ambiente em que ocorre o processo de seleção e avaliação.

Assim, no presente trabalho, adotou-se o conceito de informação proposto por Smit e Barreto (2002, p. 21-22), que a caracteriza como "estruturas simbolicamente significantes, codificadas de forma socialmente decodificável e registradas". Tal conceito é especialmente relevante para o contexto de ambientes informacionais digitais, que, similarmente, buscam armazenar informações para uso futuro, para um determinado usuário.

## **2.2 Usuários da Informação, sujeitos informacionais: características gerais**

De acordo com Sanz Casado (1994, p. 19), o usuário da informação é o "indivíduo que necessita de informação para o desenvolvimento de suas atividades". As necessidades informacionais frequentemente surgem quando um usuário se depara com um problema que demanda algum tipo de conhecimento ou orientação. Nesse contexto, o conhecimento refere-se a informações específicas, dados ou habilidades que o usuário precisa adquirir para resolver um problema ou tomar uma decisão. A orientação, por outro lado, envolve um guia ou um conjunto de diretrizes que podem ajudar o usuário a navegar partindo de um problema ou situação complexa.

Tais necessidades informacionais não são estáticas e mudam com base em diversos fatores, incluindo tarefas organizacionais como planejamento e tomada de decisões, bem como aspectos da personalidade do usuário. Além disso, à medida que as circunstâncias, interesses e objetivos dos usuários mudam, suas necessidades de informação se adaptam de forma correspondente.

Inicialmente, o indivíduo pode encontrar-se em um "estado anômalo de conhecimento", caracterizado por uma sensação de desconforto ou inadequação em relação ao seu nível atual de entendimento sobre um determinado tópico (BELKIN, 1980). O referido estado serve como um catalisador para a busca de informações adicionais. À medida que o indivíduo avalia a relevância do problema, o mesmo começa a identificar as lacunas específicas em seu conhecimento que necessitam ser preenchidas. No entanto, é importante notar que a consciência dessa necessidade informacional não leva automaticamente a uma busca ativa por informações. A decisão de prosseguir com a busca é influenciada por múltiplos fatores, incluindo a importância percebida do problema, o conhecimento prévio sobre o tema e o esforço estimado para realizar a busca (CHOO, 2003).

Em adição, Wilson (2006) complementou o conceito de necessidade informacional como uma experiência subjetiva que ocorre exclusivamente na mente de cada indivíduo, sendo inacessível diretamente ao observador. Para identificar esta necessidade, é necessário deduzir o problema por meio da observação do comportamento do indivíduo ou através da própria manifestação da necessidade pelo próprio indivíduo. Os estudos de Dervin (1992) destacaram o caráter cognitivo e não observável das necessidades informacionais, apontando para a existência de lacunas que podem gerar descontinuidade no conhecimento humano. Quando percebidas pelo indivíduo, estas lacunas suscitam o surgimento da necessidade informacional.

Belkin (1980) explicou o fenômeno da lacuna informacional a partir do conceito de estado anômalo do conhecimento. Belkin (1980, p. 44, tradução nossa) considera que, quando “o estado de conhecimento do usuário em relação a um tópico é de alguma forma inadequado em relação à capacidade da pessoa de alcançar algum objetivo”<sup>6</sup>, este pode ser entendido como anômalo. Tal percepção de lacuna de conhecimento leva o usuário a buscar ativamente por informações adicionais, a fim de corrigir essa lacuna e adquirir um nível de conhecimento mais completo e sólido. O usuário reconhece que precisa preencher esta falta de informação para tomar decisões informadas e alcançar seus objetivos. Em adição a isso, Le Coadic (2004, p. 8) destacou que

Nosso estado (ou nossos estados) de conhecimento a respeito de determinado assunto, em determinado momento, são representados por uma estrutura de conceitos ligados por suas relações: nossa imagem de mundo. Quando constatamos uma deficiência ou anomalia desse(s) estado(s) de conhecimento, encontramos um estado anômalo de conhecimento. Tentamos obter uma informação ou informações que corrigirão essa anomalia. Disso, resultará um estado novo de conhecimento (LE COADIC, 2004, p.8).

Le Coadic (2004, p. 40) ainda complementou, indagando:

[...] o que leva uma pessoa a procurar informação? A existência de um problema a resolver, de um objetivo a atingir e a constatação de um estado anômalo de conhecimento, insuficiente e inadequado (LE COADIC, 2004, p.40).

Nota-se que a ausência de conhecimento sobre um determinado assunto ou um problema específico do usuário desencadeia o processo de busca por

---

<sup>6</sup> *“the user’s state of knowledge with respect to a topic is in some way inadequate with respect to the person’s ability to achieve some goal”* (BELKIN, 1980, p. 4).

informações. Essa demanda por conhecimento exerce influência sobre o comportamento do usuário diante de seu problema informacional.

Le Coadic (2004, p. 39) sugeriu que as necessidades e os usos da informação são “interdependentes, se influenciam reciprocamente de uma maneira complexa que determinará o comportamento de um usuário e suas práticas”. Assim, considerando o contexto em que a necessidade informacional se manifesta, as habilidades e competências do usuário, bem como, a disponibilidade de acesso a recursos informacionais, tem-se o que se denomina de comportamento informacional.

Ademais, o comportamento informacional se refere às atividades, às estratégias e aos processos que os indivíduos empregam para buscar, acessar, avaliar e utilizar informações que atendam às suas necessidades informacionais (WILSON, 2000). Isso inclui a busca ativa e seletiva de fontes de informação, com o objetivo de encontrar dados que se adequem às suas necessidades específicas.

Por conseguinte, é necessário avaliar a relevância das informações recuperadas, determinando se elas são aplicáveis e úteis para a situação em questão. A confiabilidade das fontes também deve ser considerada, garantindo que as informações sejam precisas e provenientes de fontes confiáveis. Após a busca e avaliação, as informações são utilizadas para resolver problemas ou apoiar processos decisórios, fornecendo subsídios e embasamento para a tomada de ações fundamentadas e eficazes. Logo, o comportamento informacional abrange um ciclo contínuo de busca, avaliação e uso das informações para atender às necessidades informacionais dos indivíduos.

Estudos sobre comportamento informacional fornecem compreensões sobre como indivíduos interagem com a informação e como essas interações afetam suas vidas pessoais e profissionais. A este respeito, Guinchat e Menou (1994) destacaram que

[...] o comportamento com relação à informação manifesta-se nas relações do usuário com as unidades de informação, com seus produtos e serviços em diversos níveis. Que conhecimento o usuário tem da informação? Como o usuário seleciona as suas fontes? Como ele formula suas questões? Como escolhe suas informações? (GUINCHAT; MENO, 1994, p. 485).

Tal atividade é importante para a compreensão da forma como usuários utilizam a informação, como são impactados por ela e vice e versa. Sobre o estudo do

comportamento informacional, Oliveira e Ramalho (2012, p. 3) enfatizaram que o usuário tem como objetivo compreender

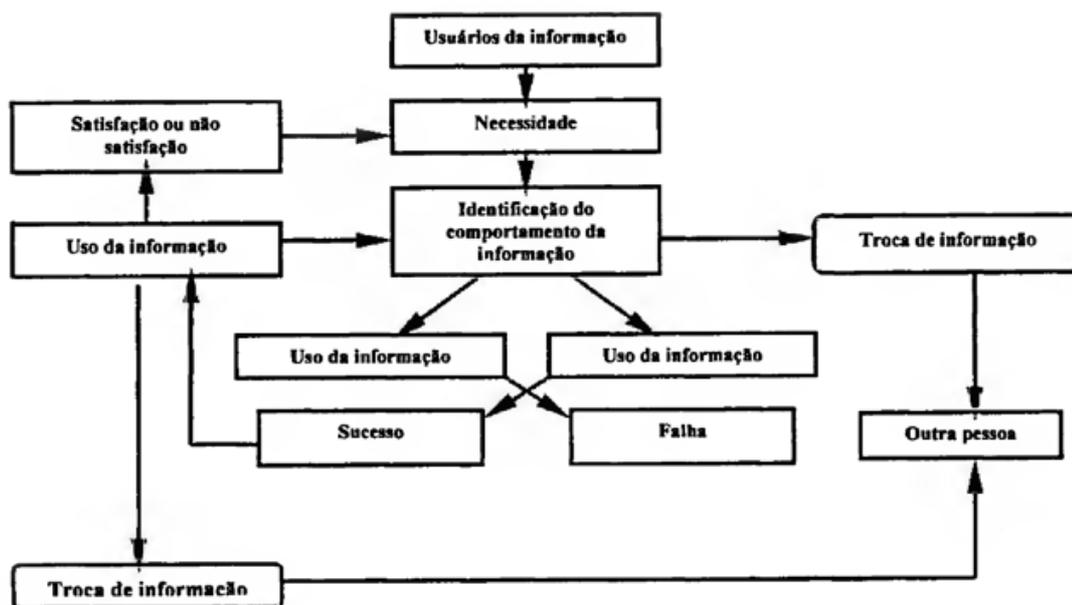
[...] porque, como, e para quais finalidades os indivíduos usam a informação, e por quais fatores esses usos são afetados. Nesse aspecto, os estudos de usuários se tornam verdadeiros canais de comunicação que se abrem entre as unidades de informação e as comunidades as quais elas servem (OLIVEIRA; RAMALHO, 2012, p. 3).

Anteriormente, Figueiredo (1994, p. 7), destacou que estudos de usuários são

[...] investigações que se fazem para saber o que os indivíduos precisam em matéria de informação, ou então, para saber se as necessidades de informação estão sendo satisfeitas de maneira adequada (FIGUEIREDO, 1994, p.7).

Wilson (2000) apresentou uma estrutura para o estudo de usuários alicerçada na questão do comportamento informacional. Tal estrutura (Figura 2) é complementar ao apresentado por Oliveira e Ramalho (2012) e por Figueiredo (1994), demonstrando a forma como um usuário parte de uma necessidade informacional e passa pela escolha da fonte de informação, avaliação da fonte, informação recuperada e avaliação de satisfação da informação recuperada.

**Figura 2 – Estrutura do Comportamento Informacional conforme Wilson (2000)**



Fonte: Wilson (2000 *apud* Araújo Júnior, 2005).

Nota-se, a partir do que é apresentado por Wilson (2006), que usuários da informação podem ou não recorrer a instituições ou sistemas de informação para satisfazer suas necessidades informacionais. O autor destacou que o usuário pode buscar informações em sistemas formais e informais, incluindo a “troca interpessoal de informação”, ou seja, buscando informações com outras pessoas. Isso é importante, pois, usuários irão buscar informações em ambientes informacionais que entendam satisfazer suas necessidades informacionais.

Em contraponto a Wilson (2006), Savolainen (2007) avaliou o conceito de comportamento informacional, argumentando que sua definição nos contextos de busca, de recuperação, de compartilhamento, de organização e de gestão da informação, é pouco clara. No entanto, Savolainen (2007) ressaltou que o principal objetivo é destacar positivamente as práticas de informação, que envolvem ações específicas de informação (RABELLO; GONZALEZ DE GÓMEZ, 2017). Nesse sentido, Savolainen (2007) se diferencia da abordagem de Wilson (2000), definindo a prática de informação como um modo de comportamento (RABELLO; GONZALEZ DE GÓMEZ, 2017).

Na perspectiva das práticas informacionais, não existe rigidez em termos de normas e estruturas exteriores e independentes, mas sim, indivíduos que continuamente atualizam as regras e os modelos (ARAÚJO, 2017). Este conceito está intimamente ligado a outro conceito, o de *accountability*, que se refere à maneira reflexiva como os indivíduos tornam disponíveis e relatáveis suas experiências e ações, tanto para si mesmos quanto para os outros (ARAÚJO, 2017).

De fato, os estudos sobre o comportamento informacional se alicerçam, de forma geral, nos processos que levam um usuário desde a busca até a avaliação, em termos de satisfação da informação recuperada. Por vezes, estes estudos desconsideram especificidades dos processos de busca e recuperação da informação. No entanto, enquanto o estudo das práticas informacionais pode fornecer entendimentos para o desenvolvimento de métodos e técnicas de busca e recuperação de informação, pode não considerar fatores subjetivos, como as emoções ou percepções individuais do usuário, que também podem influenciar sua interação com os sistemas de informação. (ARAÚJO, 2017).

Em complemento às práticas informacionais, uma segunda abordagem nos estudos de usuários privilegia na "Ação de Informação", um conceito que abrange as atividades e decisões tomadas por um usuário ao interagir com sistemas e fontes de

informação. Esta abordagem é influenciada pela teoria da ação comunicativa de Habermas (1989) e foi inicialmente proposta por Wersig e Windel (1985). Tal abordagem se dedica a explorar como os aspectos da vida cotidiana do usuário influenciam suas necessidades informacionais. Diferentemente das práticas informacionais, as quais podem ser mais estáticas ou dinâmicas, a "Ação de Informação" incorpora a subjetividade da ação humana, oferecendo uma visão mais completa das necessidades e comportamentos informacionais. Sobre tal teoria, Rabello (2017, p. 124) afirmou que o

[...] conceito de ação de informação de Wersig e Windel (1985), segundo os próprios autores, foi originalmente proposto como uma alternativa à abordagem comportamental normativa que desconsidera a relação do contexto subjetivo com os fatos da vida cotidiana, bem como com os fatores sociais, econômicos e culturais. A ação, nesse caso, seria instrumentalizada para a resolução de problemas, ou seja, para a solução de situações de necessidade de informação, considerando aspectos sociais, mas, sobretudo, subjetivos (RABELLO, 2017, p. 124).

Ao levar em consideração a dimensão subjetiva do processo informacional que envolve o usuário, expandindo a abordagem para incluir a ação informacional, emergem elementos não abordados pelo modelo de Wilson (2006). Esses elementos são de natureza social e subjetiva, conforme apontado por Rabello (2017), sendo historicamente relacionados ao progresso coletivo e à tecnologia, e proporcionam novas oportunidades para os estudos sobre usuários. Tais oportunidades podem ser integradas para melhorar os novos modelos de ambientes informacionais digitais. Nesse cenário, há uma mudança observável no papel do usuário.

Rabello (2013) apontou, complementando o apresentado por Wersig e Windel (1985) e Wilson (2006), que o usuário da informação, atualmente, não age apenas como receptor da informação, mas também como produtor e disseminador da informação. Compreendendo que o usuário detém uma forma de agência social no contexto do uso e da produção de informações, conforme descrito por Weber (1977), configura-se como um sujeito ativo no campo informacional. Sobre Isso, Rabello (2013) expôs que,

Nessa direção, o sujeito pode se referir não apenas àquele indivíduo que gerencia ou faz uso da informação num sistema, mas também àquele que valida e produz informações e conhecimentos, apropria-se

dele e o dissemina em diferentes formas de interação e contextos (RABELLO, 2013, p. 58).

Embora tal fenômeno não seja novo, o seu crescimento e alcance foram impulsionados pelas TDICs. O desenvolvimento das TDICs permitiu o acesso a ferramentas digitais e online capazes de agregar e compartilhar uma quantidade de dados que, sem elas, não seria possível (MARQUESONE, 2016). Como resultado, os repositórios de dados se expandiram e passaram a receber não apenas dados e informações selecionados, mas, também, informações provenientes de diversas fontes, produzidas por diversos perfis de sujeitos informacionais. Rabello (2013) destacou que o sujeito informacional

[...] deixa de ser concebido estritamente como um sujeito passivo, um reservatório (container) de informação, bem como deixa de ser interpretado somente como um objeto para o escrutínio de práticas e comportamentos por intermédio do estudo de processos cognitivos particulares. Entra em cena, portanto, a noção unívoca de usuário-produtor de conhecimento (RABELLO, 2013b, p. 182).

O sujeito informacional desempenha o papel de produtor e usuário da informação, por meio de ações informacionais que visam atender seus objetivos. Tais ações ocorrem em diversos espaços e contextos. Logo, a compreensão desses contextos específicos pode ser utilizada para adequar as necessidades informacionais dos usuários e as recuperações de informação realizadas por um SRI.

No presente trabalho, o conceito de "sujeito da informação" assume um papel central e se alinha profundamente com as reflexões pós-positivistas de Gonzalez de Gómez (2017). A indissociabilidade do sujeito e seu contexto é postulada como um elemento-chave, guiando as investigações e análises empíricas ao longo deste trabalho.

O sujeito da informação é aqui compreendido não como uma entidade isolada, mas como um ator social inserido em um complexo sistema de estruturas de significado e relações sociais. A tese endossa o entendimento de que este sujeito informacional é profundamente moldado e influenciado pelo seu contexto, que por sua vez é também ativamente construído e modulado pelas interações sociais em que o sujeito está imerso. Tal inter-relação é destacada como uma alavanca crítica para entender as necessidades informacionais específicas de cada sujeito, tornando o contexto uma variável indissociável no estudo do comportamento informacional.

O paradigma centrado no usuário, que serve como base para esta pesquisa, afasta-se da mera interação com sistemas de informação para abraçar um entendimento mais profundo do sujeito da informação e suas necessidades em um ambiente específico. O valor inerente desta abordagem é que ela reconhece e integra a complexidade do sujeito informacional, considerando não apenas suas necessidades cognitivas, mas também as influências sociais e contextuais que atuam sobre ele. Certamente, aqui está a reescrita sem utilizar a primeira ou terceira pessoa:

Neste trabalho, o usuário é concebido como um sujeito imerso em uma condição contextual única e específica. A abordagem adotada para o desenvolvimento do modelo proposto é fundamentada nessa compreensão multidimensional do usuário. Reconhece-se que cada sujeito informacional opera dentro de um conjunto de circunstâncias e relações que são ativas, dinâmicas e mutuamente constitutivas. Esse enriquecimento da análise ressoa com as contribuições de Gonzalez de Gómez (2017), que também enfatizou a indissociabilidade entre o sujeito informacional e seu contexto.

Ao lidar com SIs digitais, é essencial reconhecer a diversidade de contextos em que esses sistemas operam. Nesse ambiente multifacetado, a não identificação do potencial usuário se torna um desafio significativo, uma vez que os sistemas digitais funcionam em uma ampla gama de cenários de produção, com diversos usos possíveis. Levar em consideração esses contextos emerge como uma estratégia valiosa para resolver o problema da não identificação dos usuários e aumentar as chances de os sistemas atenderem às suas demandas.

Assim, é possível enfrentar a rigidez imposta por estes sistemas, proporcionando resultados mais pertinentes e personalizados para cada usuário. Portanto, é relevante analisar o funcionamento dos SRIs e dos SIs em ambientes digitais, a fim de compreender como a utilização destes contextos específicos pode contribuir para o aprimoramento desses sistemas.

### **2.3 Sistemas de informação e recuperação da informação**

Ao longo do tempo, a definição de sistema de informação (SI) passou por transformações devido ao avanço tecnológico e às ferramentas de comunicação. Embora alguns marcos históricos semelhantes possam ser observados, é importante

distinguir as características específicas destes sistemas, considerando que os processos que envolvem os fluxos informacionais estão em constante mudança.

Cesarino (1978, p. 224) indicou que um SI é “uma organização ou unidade social que procura atingir um objetivo específico, servindo, na realidade, de fonte intermediária entre o produtor e o consumidor da informação”. Cohen (1995, p. 14) complementa tal afirmação onde pontuou que um SI é um “conjunto de canais formais e informais de comunicação da informação dentro de uma organização ou de uma comunidade”. Nota-se que, neste contexto, um SI é aquele que atua como ferramenta comunicacional de informação para uma comunidade, com um objetivo específico, de produtor para usuário.

Araújo (1995, p. 1) reforçou o mesmo posicionamento dizendo que SIs

[...] são aqueles que, de maneira genérica, objetivam a realização de processos de comunicação. Alguns autores contextualizam sistemas de informação mais amplamente para incluir sistemas de comunicação de massa, redes de comunicação de dados e mensagens etc., independentemente da forma, natureza ou conteúdo desses dados e mensagens (ARAÚJO, 1995, p.1).

Tais sistemas pressupõem a existência de um arcabouço informacional que se adapta as mudanças da sociedade para as quais são pensados, sendo chamados de "estoques de informação" (SMIT; BARRETO, 2002). Os SIs são projetados para lidar com a movimentação de tais estoques, que armazenam registros feitos pelos usuários para uso presente ou futuro. De um modo geral, os estoques informacionais são recursos que armazenam informações e dados em diferentes formatos. Tais estoques são importantes para a preservação, acesso e compartilhamento do conhecimento, contribuindo para o funcionamento dos sistemas de informação. Sobre isso, Smit e Barreto (2002, p. 14) apontaram que a

[...] produção da informação documentária é operacionalizada por meio de práticas bem definidas e se apoia em um processo de transformação orientado por uma racionalidade técnica específica; representa atividades relacionadas à reunião, seleção, codificação, redução, classificação e armazenamento de informação. Todas essas atividades estão orientadas para a organização de estoques de informação, de uso imediato ou futuro. Esse repositório de informação representa um estoque potencial de conhecimento e é imprescindível que exista, para que se realize a transferência de informação (SMIT; BARRETO, 2002, p. 4).

Embora as informações armazenadas em bancos de dados, bibliotecas, arquivos ou museus, tenham o potencial de gerar conhecimento, isso só ocorre por

meio de uma ação de comunicação mutuamente acordada entre a fonte (os estoques) e o receptor (SMIT; BARRETO, 2002). Assim, os estoques de informação dependem de processos que visam à disponibilização da informação produzida e armazenada para os indivíduos, a partir de um fluxo que conecta o estoque e o indivíduo.

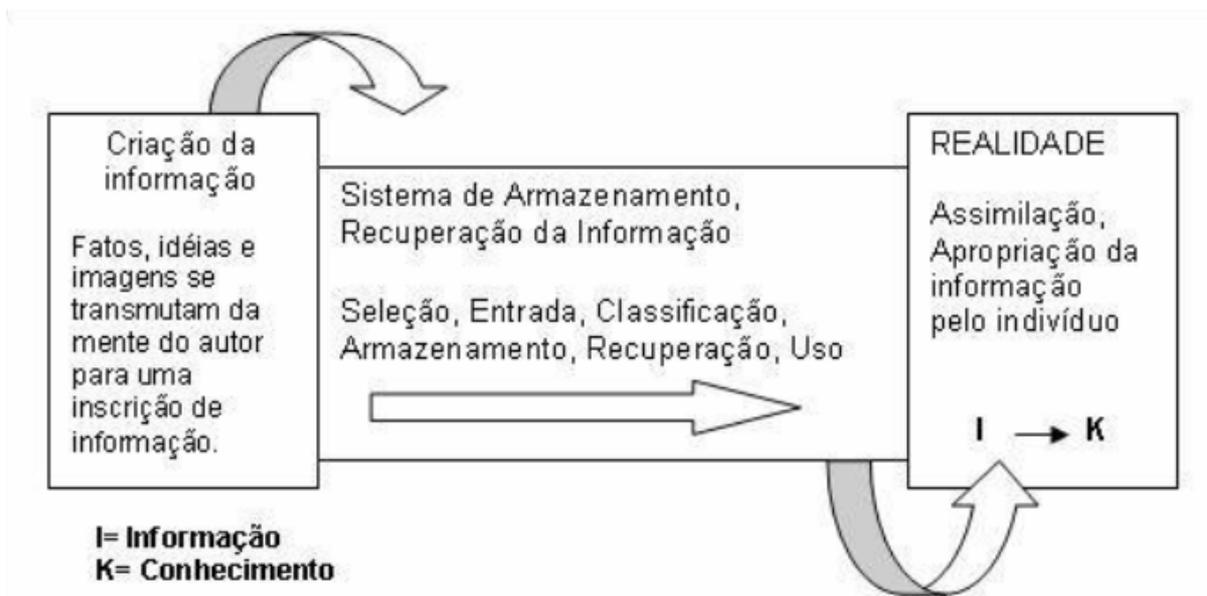
Sobre os fluxos informacionais destes estoques, Smit e Barreto (2002) destacaram que

[...] dois critérios permeiam o fluxo da informação entre os estoques, ou espaços de informação, e os usuários: o critério da tecnologia da informação, que almeja possibilitar o maior e melhor acesso à informação disponível, e o critério da Ciência da Informação, que intervém para qualificar este acesso em termos das competências que o receptor da informação deve ter para assimilar a informação, ou seja, para elaborar a informação para seu uso, seu desenvolvimento pessoal e dos seus espaços de convivência. Não é suficiente que a mensagem esteja disponível, ela deve também poder ser apropriada pelo receptor (SMIT; BARRETO, 200, p. 15-16).

Logo, as regras de seleção do que será armazenado em um estoque informacional devem levar em conta as especificidades dos usuários deste estoque, cabendo à instituição responsável pelo estoque determinar o que armazenar e como apresentar tais informações, através de sistema próprio de informação. É importante indicar que a “[...] produção dos estoques de informação não possui um compromisso direto e final com a produção do conhecimento” (BARRETO, 1994, p. 4).

Os fluxos informacionais esperados de um SI clássico, ou seja, um SI projetado com um usuário específico em vista, se atentam para o observado na Figura 3. A informação é produzida em contextos diversos e selecionada com base no perfil e nas necessidades de um potencial usuário. É armazenada levando-se em conta critérios específicos para atender a necessidade informacional de tal potencial usuário (SMIT; BARRETO, 2002). Após buscar e recuperar a informação, o usuário avalia sua relevância e, se possível, a emprega para resolver o problema ou atender à necessidade que motivou sua busca inicial. Em alguns casos, sistemas de informação digitais podem adotar características desse fluxo informacional quando são projetados para consultas específicas e direcionadas.

**Figura 3 - Fluxo interno e os fluxos externos da informação**



Fonte: Smit e Barreto (2002, p. 14).

A partir desse fluxo informacional, sistemas de informação podem ser classificados de acordo com sua função, área de aplicação ou modelo de processamento de informações. Alguns tipos de sistemas de informação são: sistemas de processamento de transações, sistemas de informação gerencial, sistemas de suporte de decisão, sistemas de informação executiva, sistemas de recuperação de informações, sistemas de informação geográfica, sistemas de automação de escritório, sistemas de comércio eletrônico, sistemas de informação em saúde, entre outros (BIO, 1996). Cada um destes sistemas tem sua própria finalidade e é usado para atender a diferentes necessidades das organizações e usuários.

Pode-se classificar este conjunto de SIs em digitais e físicos. Enquanto os sistemas físicos utilizam formatos de armazenamento tangíveis, como livros e revistas, os sistemas digitais utilizam arquivos eletrônicos, como bancos de dados e imagens digitais. A acessibilidade da informação, ou seja, a capacidade de um usuário de compreender a nível de forma e conteúdo a informação que lhe é recuperada, é outra diferença importante, pois os sistemas tradicionais só podem ser acessados em locais físicos, enquanto os digitais podem ser acessados de qualquer lugar do mundo com conexão à internet. Sobre SRIs digitais, Vickery e Vickery (2004) apresentaram seguinte definição:

[...] a essência da recuperação eletrônica é que uma coleção de mensagens é armazenada em algum meio legível por computador [...] e é acessada por um software executado em um computador ao qual

o armazenamento está vinculado. Um sistema pode ser pessoal [...], institucional [...], ou público (VICKERY; VICKERY, 2004, p. 117, tradução nossa).<sup>7</sup>

Além disso, os sistemas digitais possuem uma capacidade de armazenamento muito maior em comparação com os sistemas físicos, permitindo o armazenamento de um número virtualmente infinito de documentos, limitado à sua capacidade física e de escalonamento, ou expansão, do banco de dados (BD). Além disso, o método de consulta aos estoques informacionais desempenha um papel importante na eficácia do SRI. Nos ambientes digitais, as buscas podem ser mais ágeis e exatas devido à utilização de recursos tecnológicos como índices e palavras-chave. Em contrapartida, nos sistemas tradicionais, a morosidade e o esforço manual são mais notáveis, não necessariamente devido a problemas na indexação, mas sim, à demanda de trabalho físico implicada no processo de recuperação da informação. Por fim, a interatividade é um aspecto que diferencia significativamente os sistemas digitais, uma vez que permitem que os usuários interajam com os documentos de várias maneiras, enquanto nos sistemas tradicionais essa interatividade é limitada.

Neste trabalho, busca-se compreender como a incorporação de atributos contextuais de uso da informação podem melhorar os sistemas de recuperação de informações digitais. Para isso, é necessário analisar as particularidades dos sistemas de recuperação de informações, especialmente os sistemas de recuperação de informações digitais. Portanto, daremos prioridade a este tipo de SRI, e as descrições fornecidas serão específicas para este modelo de funcionamento do SI.

Um SRI digital é um tipo específico de SI, com uma finalidade bem definida. Enquanto os sistemas de informação, em geral, visam armazenar, processar e disseminar informações de forma geral e estruturada, para que os usuários possam acessá-las e utilizá-las eficientemente, os sistemas de recuperação da informação digital têm como objetivo recuperar informações específicas dentro de um conjunto de

---

<sup>7</sup> “[...] the essence of electronic retrieval is that a collection of messages is stored in some computer-readable medium [...] and is accessed by software run on a computer to which the store is linked. A system may be personal [...], or institutional [...] or public” (VICKERY; VICKERY, 2004, p. 117).

dados ou documentos, usando técnicas de busca e recuperação da informação. Araújo Júnior (2005, p. 69) descreve a busca e a recuperação da informação

[...] como o processo de localizar documentos e itens de informação que tenham sido objeto de armazenamento, com a finalidade de permitir o acesso dos usuários aos itens de informação, objetos de uma solicitação (ARAÚJO JÚNIOR, 2005, p. 69).

De forma direta, sistemas de recuperação da informação digital podem ser definidos como

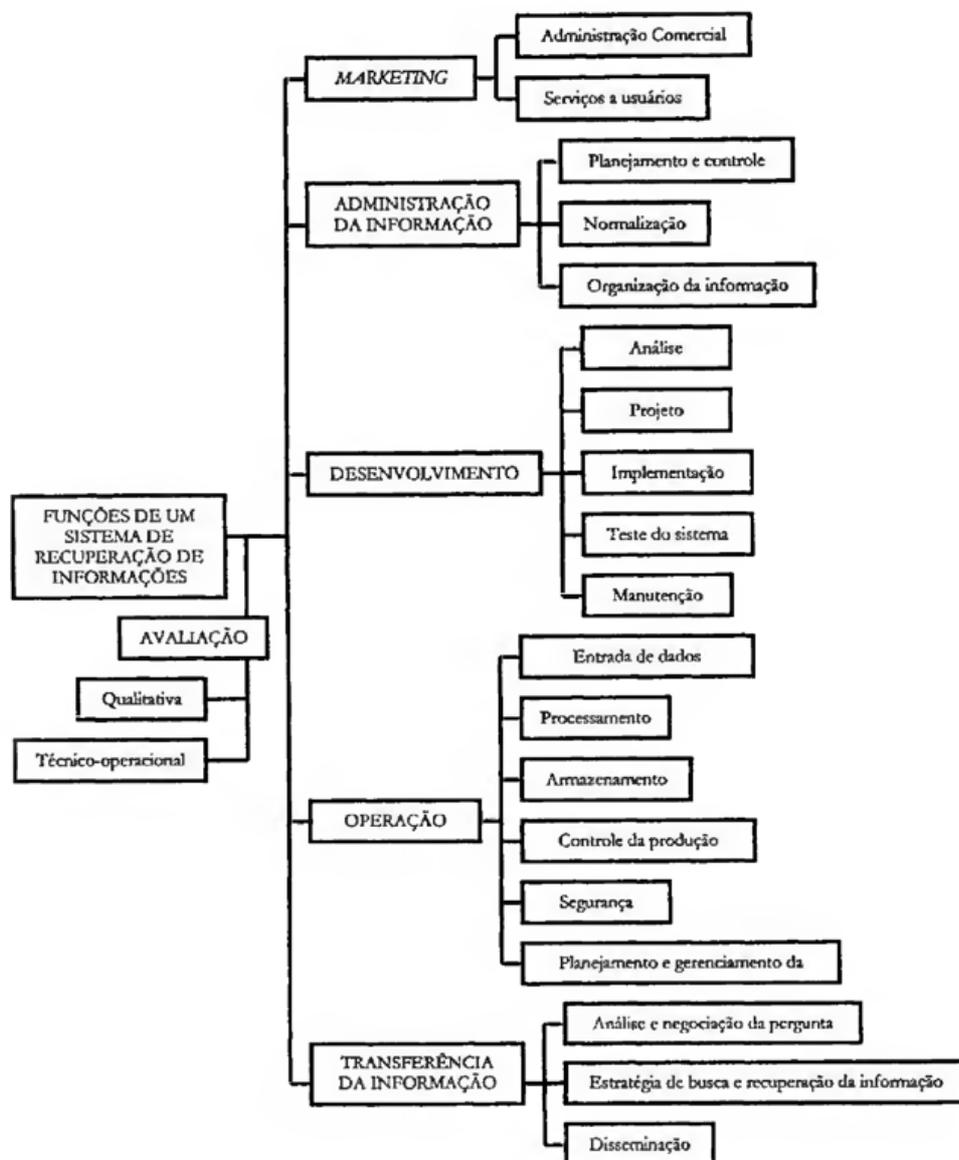
[...] um sistema de operações interligadas para identificar, dentre um grande conjunto de informações (uma base de dados, por exemplo), aquelas que são de fato úteis, ou seja, que estão de acordo com a demanda expressa pelo usuário (ARAÚJO JÚNIOR, 2005, p. 77).

Como visto, em consonância com o apresentado sobre SIs, as preocupações com as necessidades dos usuários são mantidas, bem como, as preocupações relativas ao processamento das informações estocadas. Ao projetar e desenvolver um SRI digital, é importante levar em consideração as necessidades e especificidades dos usuários potenciais. A função principal dos sistemas de recuperação da informação digital é satisfazer a necessidade de informação do usuário, levando-o ao documento correto, ou informação que irá atender sua necessidade específica de informação (ARAÚJO JÚNIOR, 2005).

É possível, assim, descrever que os sistemas de recuperação de informação digital têm a função de representar o conteúdo dos documentos presentes no seu estoque informacional e disponibilizá-los para o usuário, de forma que possibilite a seleção rápida dos itens que atendam total ou parcialmente às suas necessidades de informação, expressas através de uma consulta de busca (FERNEDA, 2003).

A Figura 4, apresentada por Araújo Júnior (2005, p. 82), apresenta algumas funções possíveis para sistemas de recuperação de informações, digitais ou não, considerando as necessidades informacionais de usuários.

Figura 4 - Funções de um sistema de recuperação de informações



Fonte: Araújo Júnior (2005, p. 82).

Sobre os tipos de SRIs aptos a estas funções, no Quadro 2, Rowley (2002, p. 72) destacou os diferentes tipos de SRIs.

Quadro 2 - Tipos de SRIs organizados por suporte

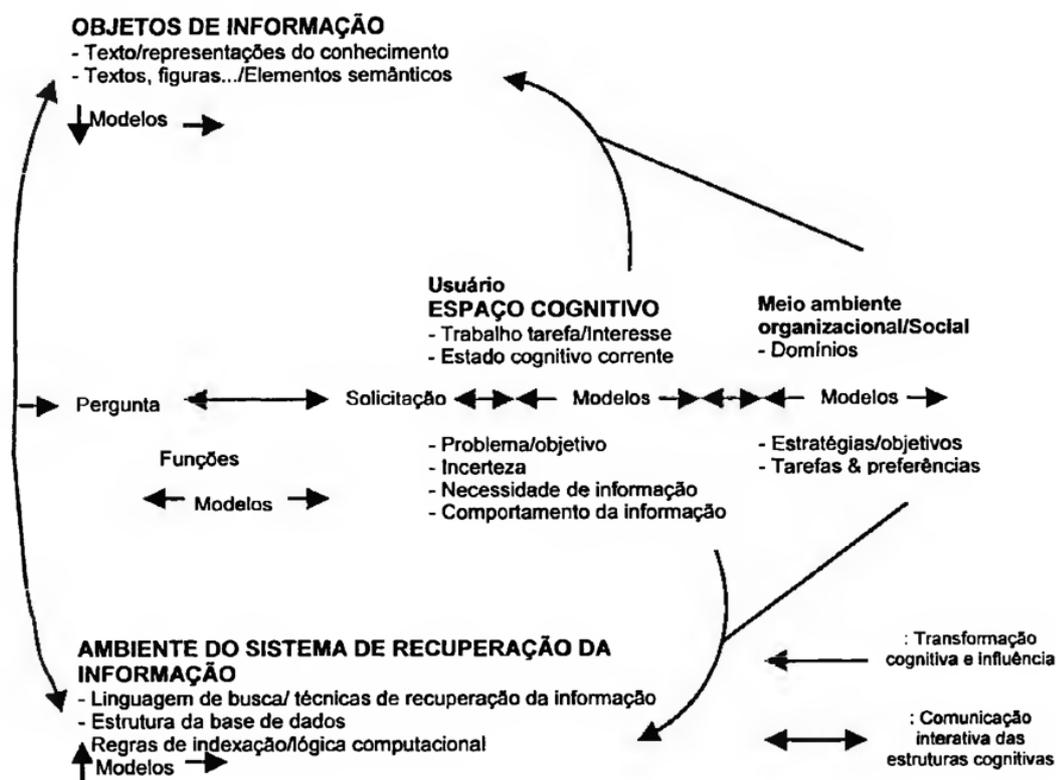
	<i>Características dos usuários</i>	<i>Ambiente</i>	<i>Tarefas</i>	<i>Tecnologia</i>
<i>Serviços de busca em linha</i>	Usuários especialistas e gerentes de informação	Escritório, biblioteca, universitária, centro de informação empresarial	Recuperação de informações, importação de informações e integração com outros documentos	Variedade de estações de trabalho; configurações antigas com ligação direta com o serviço; mais aplicações na ponta; vínculos na Internet
<i>Cederrom</i>	Depende da base de dados – pode incluir crianças, usuários em geral de bibliotecas públicas, usuários especializados e outros	Biblioteca, aeroporto, residência, escritório	Recuperação de informações, importação de informações e integração com outros documentos	Freqüentemente multimídia, interface gráfica, mouse
<i>Internet</i>	Internautas – predomínio de docentes, estudantes e pessoas do sexo masculino	Lugar de estudo/trabalho, residência	Comunicação por correio eletrônico, vendas, transferência de arquivos	Microcomputadores de mesa e portáteis, com teclado, monitor e mouse
<i>Catálogos em linha de acesso público</i>	Usuários de biblioteca – o perfil depende do tipo de biblioteca	Biblioteca, escritório/residência, outros locais públicos	Estritamente definidas – identificação da disponibilidade de livros, busca de informação	Às vezes telas grandes, telas sensíveis ao toque, teclados especiais, mas também com acesso por meio de equipamento comum de escritório; acesso remoto e local; podem usar estações de trabalho diferentes
<i>Sistemas de gerenciamento de documentos</i>	Usuários institucionais com alguma experiência em comum do sistema, e objetivos e tarefas comuns	De escritório, mas pode também estender-se para operação móvel e uso em unidades de produção em, p.ex., trens e carros	Consulta a arquivo da empresa na execução de tarefas ligadas ao trabalho	Estações de trabalho ligadas a uma poderosa central de processamento; algumas aplicações serão de ponta

Fonte: Rowley (2002 *apud* Araújo Júnior, 2005, p. 72).

A fim de recuperar, para o usuário, uma informação ou conjunto de informações potencialmente relevantes, diversas estratégias e modelos de operação podem ser considerados para o projeto e operação de sistemas de recuperação de informações digitais, a partir destes tipos apresentados por Rowley (2002). Explorar estes modelos pode dar luz às problemáticas observadas em ambientes de informação digitais.

Em um estudo anterior, Ingwersen (1996) introduziu um modelo de processos de RI em ambientes computacionais. O modelo enfatizou o espaço cognitivo do usuário da informação, o qual pode ser entendido como o ambiente onde a informação é processada, interpretada e aplicada. Tal contexto frequentemente influencia como os indivíduos buscam, avaliavam e empregam informações, conforme ilustrado na Figura 5.

Figura 5 - Modelo de SRI proposto por Ingwersen (1996)



Fonte: Ingwersen (1996 *apud* Araújo Júnior, 2005).

A abordagem geral em relação à busca de informação em um sistema de RI, para Ingwersen (1996), é focada na busca ativa, na qual o usuário aborda o SRI digital, o que é comum em muitos modelos de busca de informação. No entanto, Ingwersen (1996) identifica uma série de outros elementos que devem ser considerados. Primeiro, Ingwersen (1996) destaca que cada área do modelo (que é um modelo de comportamento de busca de informação) inclui várias entidades, como o usuário da informação, o autor do documento, o intermediário, a interface e o sistema de RI.

Cada uma destas entidades tem uma função específica na interação do usuário com o sistema de RI, e cada função é baseada em modelos cognitivos explícitos ou implícitos do domínio da busca de informação. Desta forma, todos os elementos de um SRI devem considerar as especificidades de necessidade informacional de seus usuários, bem como, suas condições ambientais e sociais.

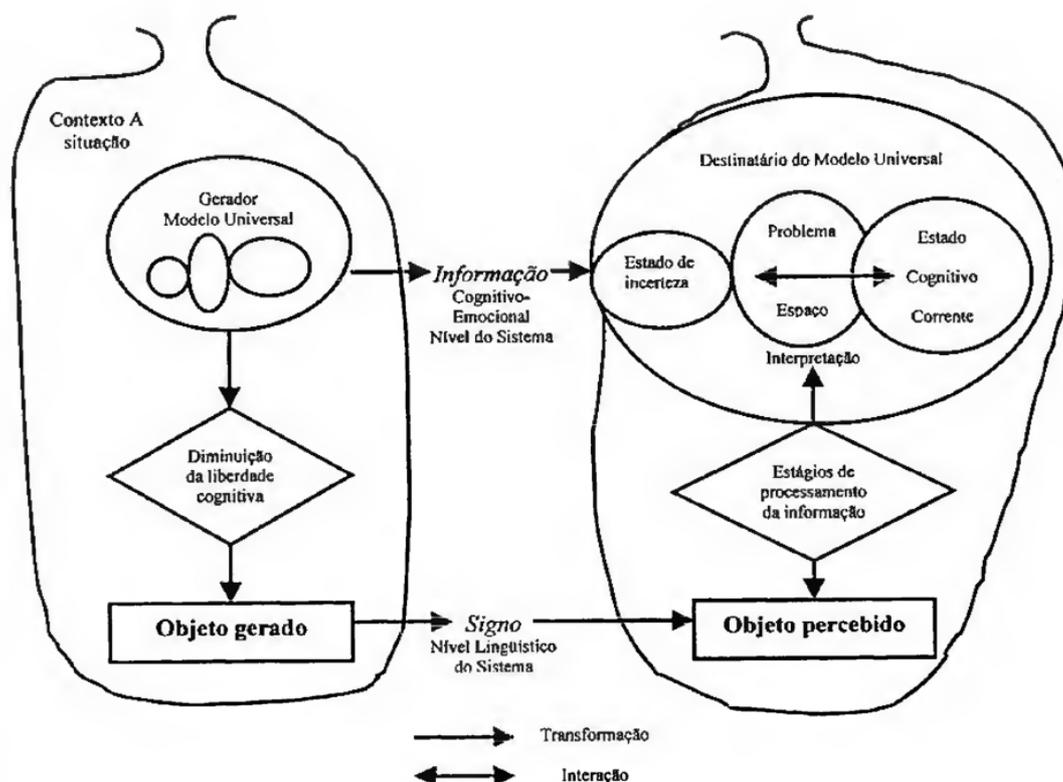
Ingwersen (1996) enfatizou a importância de incluir o SRI como parte do modelo abrangente de comportamento de busca de informações, entendendo a recorrência deste processo e a importância de sua relação com a realidade e o ambiente do usuário. Isso é importante, pois, seu modelo considera que existem

transformações cognitivas que ocorrem entre o cotidiano dos usuários e a pesquisa. Assim, o autor aponta para a necessidade de comunicação efetiva entre todas as entidades envolvidas no sistema de RI, em especial, aquelas que consideram a realidade do usuário.

Embora Saracevic (1996) tenha sugerido que o modelo de Ingwersen (1996) tivesse problemas para avaliação de uso, em especial, no que tange a análise de atualização à longo prazo, o modelo de Ingwersen (1996) ainda é considerado uma contribuição importante para a compreensão do comportamento de busca de informação em sistemas de recuperação da informação.

Isso significa que o modelo é "incompleto" para o estudo da informação dentro deste SRI, pois não considera como os usuários chegam ao ponto de fazer a pesquisa, nem como suas estruturas cognitivas são afetadas pelos processos de decidir como e quando se mover em direção à pesquisa de informações. Estas questões podem ser discutidas em termos do ambiente social ou organizacional, mas não são explicitamente abordadas no modelo de Ingwersen (1996), de acordo com Wilson (2006).

Figura 6 - Modelo atualizado de SRI de Ingwersen (1999)



Fonte: Ingwersen (1999, *apud* Araújo Júnior, 2005).

O modelo proposto por Ingwersen (1999), apresentado na Figura 6, é uma versão atualizada do modelo anteriormente proposto pelo mesmo autor em 1996. Este modelo considera dois aspectos cruciais: as tarefas de trabalho e a percepção cognitiva do usuário, ambos inseridos em um contexto situacional. O contexto situacional compreende o ambiente físico do usuário, as atividades que ele está realizando, as restrições e recursos disponíveis, as características do sistema ou dispositivo que está utilizando, bem como suas necessidades e expectativas.

No contexto aqui mencionado, a percepção refere-se à maneira como o usuário interpreta e compreende as tarefas que está realizando. Isso inclui seu nível de interesse, a identificação de problemas a serem solucionados e as tarefas específicas que precisa executar. A percepção desempenha um papel fundamental na origem das necessidades de informação, conforme enfatizado por Ingwersen (1999). Do ponto de vista cognitivo, a percepção do usuário durante uma tarefa tende a ser mais estável ao longo da sessão de recuperação de informações em comparação com a necessidade de informação dinâmica correspondente, ou seja, aquela que pode evoluir durante o processo de recuperação.

A percepção da tarefa de trabalho é um elemento valioso para a recuperação de informações, pois fornece ao sistema um contexto essencial para recuperar informações relevantes para o usuário enquanto ele executa suas tarefas. Isso permite aprimorar as estratégias de busca e, conseqüentemente, melhorar o processo de recuperação de informações.

Deste modo, Ingwersen (1999) apresentou um modelo que privilegia sua atualização, no sentido de satisfazer as necessidades informacionais adaptadas aos usuários ao longo do tempo, dependendo de sua atualização informacional. De fato, esta é uma preocupação observada em SRIs, digitais ou não, em que a atualização cognitiva do usuário pode impactar no entendimento de relevância da informação recuperada.

Numa abordagem diferente daquela proposta por Ingwersen (1999), centrada em um modelo abrangente de busca de informação, Choo (2003, p. 76), construído a partir das ideias de Saracevic (1975), propôs um modelo de fluxo de trabalho aplicável, em primeira instância, às SRIs digitais.

- 1.O usuário tem um problema a resolver (características do usuário, declaração do problema).
- 2.O usuário procura resolver o problema formulando uma pergunta e iniciando uma interação com um sistema de informação (declaração da pergunta, características da pergunta).
- 3.Interação de pré-investigação com um pesquisador intermediário, humano ou computador (características do pesquisador, análise da pergunta)
- 4.Formulação de uma busca (estratégia de busca, características da busca).
- 5.Atividade de busca e interações (busca).
- 6.Entrega das respostas ao usuário (itens armazenados, formatos despachados).
- 7.Avaliação das respostas pelo usuário (relevância, utilidade). (CHOO, 2003, p. 76).

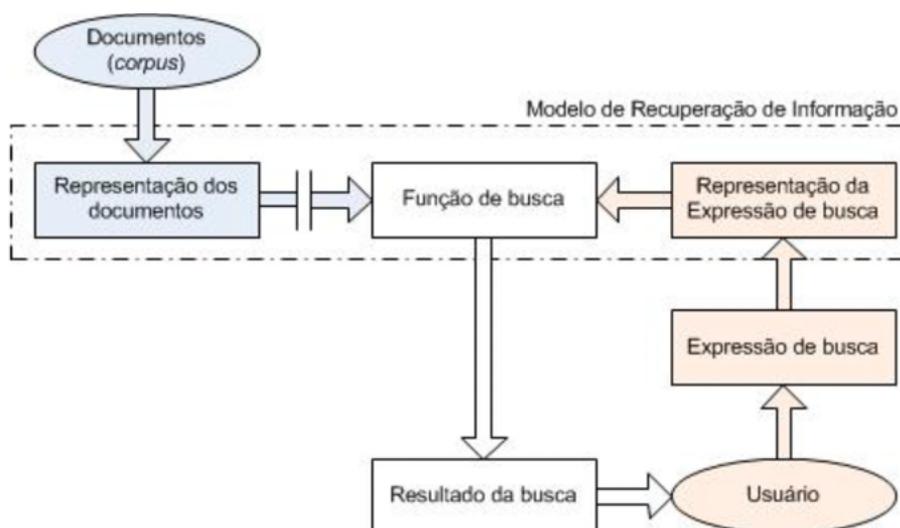
Sobre a qualidade deste modelo, Choo (2003) indicou que

A análise dos dados empíricos mostrou que "os modelos sugeridos foram aprovados, ou seja, os elementos sugeridos pelos modelos apresentaram uma significativa relação com os resultados armazenados". Confirmou-se, por exemplo, que o contexto de uma pergunta é importante, inclusive os antecedentes que levam à formulação da pergunta e a intenção por trás do uso da informação a ser armazenada. Diferentes tipos de perguntas - classificadas de acordo com sua clareza, especificidade, complexidade, etc. - terão provavelmente diferentes níveis de desempenho no que diz respeito

ao armazenamento da informação. Ciclos de busca tendem a melhorar os resultados, já que os resultados intermediários podem ser revistos e as estratégias de busca refinadas de acordo com eles (CHOO, 2003, p. 76).

Isto entra em consonância com o modelo apresentado por Ingwersen (1999), sobre a operação de SIs. Em um outro modelo de operação, de acordo com Ferneda (2012), o processo de recuperação de informações em um SRI digital, o usuário realiza uma abordagem ao sistema por meio de uma expressão de busca para acessar o SRID. O sistema, por sua vez, utiliza uma função de busca como mediação para representar os documentos presentes no corpus e, em caso de correspondência com a expressão de busca, apresenta-os como resultado para o usuário. Na Figura 7, é apresentado este processo.

**Figura 7 - Fluxo de um sistema de recuperação de informações digital conforme Ferneda (2012)**



Fonte: Ferneda (2012, p. 14).

Em um SRI, é responsabilidade do usuário formular uma expressão de busca que descreva adequadamente sua necessidade de informação (ARAÚJO JÚNIOR, 2005). A busca de informação neste sistema pode ser realizada através de linguagem natural ou palavras-chave, com o objetivo de recuperar um conjunto de documentos relevantes para o usuário.

Após apresentação dos termos de busca, o sistema os relaciona com o seu estoque a partir de representações. A representação de documentos em sistemas

digitais de recuperação de informação tem como objetivo descrever ou identificar cada documento da coleção com base em seu conteúdo. (FERNEDA, 2003). Segundo Novellino (1996),

A principal característica do processo de representação da informação é a substituição de uma entidade linguística longa e complexa - o texto do documento - por sua descrição abreviada. O uso de tal sumarização não é apenas uma consequência de restrições práticas quanto ao volume de material a ser armazenado e recuperado. Essa sumarização é desejável pois sua função é demonstrar a essência do documento. Ela funciona então como um artifício para enfatizar o que é essencial no documento considerando sua recuperação, sendo a solução ideal para organização e uso da informação (NOVELLINO, 1996, p. 38).

Tal representação é, geralmente, realizada por meio do processo de indexação (FERNEDA, 2003).

A indexação é um processo essencial de representação realizado em um SRI, no qual o indexador, seja ele humano ou um sistema digital, examina o documento ou arquivo e distingue a informação relevante daquela periférica, a fim de representá-lo de maneira adequada para posterior recuperação (LIMA; CAMPOS, 2022). Segundo Lima e Campos (2022, p. 2) “a indexação é vista como “[...] a representação do conteúdo dos documentos por meio de símbolos especiais, quer retirados do texto original, quer escolhidos numa linguagem de informação ou de indexação”. Ainda sobre indexação, Lima e Campos (2022, p. 2) indicaram que

[...] esse processo é realizado em duas etapas: a primeira é a da análise do documento para identificação de seu conteúdo informacional; a segunda, a de tradução dos conceitos nos termos de uma linguagem de indexação, utilizando-se os sistemas de organização do conhecimento, do tipo tesouros e sistema de classificação bibliográfico (LIMA; CAMPOS, 2022, p. 2).

Essa representação permite identificar o documento e definir seus pontos de acesso para a busca, podendo também ser utilizada como seu substituto. Esse processo é descrito por Ferneda (2003) como parte importante do processo de recuperação de informação, em que a representação dos documentos é essencial para a mediação entre a expressão de busca do usuário e os resultados apresentados pelo sistema.

A partir de indexação, documentos e informações são recuperados para o usuário, que avalia sua relevância e o potencial de utilização mediante sua necessidade informacional. Ocorre que a relevância de um documento pode ser

afetada pelo contexto em que o termo aparece ou pela idade do documento, por exemplo (FERNEDA, 2003). Isso representa um desafio para os processos de RI em um SRI digital no que tange sua representação e indexação.

Os modelos apresentados não esgotam as possibilidades de proposição de estratégias para o projeto de sistemas de recuperação de informações clássico ou digital, mas dão uma boa ideia das possibilidades de sua operacionalização. Podemos, a partir destes modelos, afirmar que a estratégia de busca em sistemas de recuperação de informação é baseada nas necessidades de informação dos usuários, que possuem um estado anômalo de conhecimento (BELKIN, 1980).

Além disso, os resultados de um SRI podem envolver não apenas documentos potencialmente relevantes, mas também a avaliação e o julgamento da informação pelos usuários, cujo conhecimento pode ser alterado durante a interação com o sistema (ARAÚJO JÚNIOR, 2005). Observa-se grande preocupação com a questão do alinhamento das informações documentadas no repositório e as necessidades dos usuários. Tal entendimento de SRIs enfatiza a importância da interação entre o usuário e o sistema, bem como, a necessidade de adaptar os mecanismos de recuperação de informação as necessidades e expectativas dos usuários.

É importante destacar, no entanto, que SRIs possuem dificuldades relativas às suas especificidades e estratégias apresentadas, conforme apontado por Ferneda (2003), Araújo Júnior (1995), Saracevic (1996), Ingwersen e Jäverlin (2005), entre outros autores. Dessa forma, os modelos apresentados, mesmo sendo considerados para o projeto de SRIs e SRIDs, podem possuir dificuldades em alguns aspectos, os quais são caros para sua concepção, e devem ter melhor analisados. Em especial, sistemas de recuperação da informação digital que utilizam de algoritmos computacionais para o processamento de informações e operação do sistema, alguns pontos se elevam como mais preocupantes, tais como os descritos a seguir.

Sobre isso, Ingwersen e Jäverlin (2005) apresentam algumas problemáticas relacionadas à recuperação de informação em SRI digital. São elas:

[...] Carência do usuário e das tarefas. [...] Carência de interação e requisições dinâmicas. [...] Carência de variabilidade tática. [...] Carência da expectativa de incerteza. [...] Carência de relevância orientada ao usuário. [...] Carência de variedade dos bancos de dados. [...] Premissa de independência documental e negligência de sobreposição documental. [...] Insuficiência de recordação para precisão. [...] Excesso de média nas buscas. [...] Apenas recuperação

documental (INGWERSEN; JÄVERLIN, 2005, p. 7–9, tradução nossa).<sup>8</sup>

Em complemento, de acordo com Vickery e Vickery (2004), os problemas fundamentais da recuperação da informação em ambientes digitais estão relacionados com natureza das mensagens armazenadas, como registros no sistema e a forma como essas mensagens se relacionam com as consultas que serão realizadas pelos usuários.

Isso significa que a efetividade do SRI dependerá da capacidade do sistema de compreender as necessidades de informação dos usuários e de apresentar resultados relevantes e precisos. Em outras palavras, se o sistema não for capaz de capturar as necessidades dos usuários e fornecer resultados precisos, a recuperação das informações será comprometida. Para Vickery e Vickery (2004),

[...] os problemas centrais da recuperação da informação surgem da natureza das mensagens armazenadas como registros no sistema e da relação dessas mensagens com as consultas que provavelmente serão feitas ao sistema. (VICKERY; VICKERY, 2004, p. 118, tradução nossa).<sup>9</sup>

Também, sobre isso, Vickery e Vickery (2004) expuseram que

[...] nos sistemas de recuperação de informações, os "valores" (por exemplo, os textos) armazenados são de variedade ilimitada, os termos de busca apresentados nas consultas são imprevisíveis, e a relação entre as mensagens armazenadas e as consultas processadas é frequentemente ambígua. (VICKERY; VICKERY, 2004, p. 118, tradução nossa).<sup>10</sup>

---

<sup>8</sup> “[...] Lack of users and tasks. [...] Lack of interaction and dynamic reaquests. [...] Lack of tactical variability. [...] Lack of uncertainty. [...] Lack of user-oriented relevance. [...] Lack of variety in collections. [Assuming document Independence and neglectin overlapping. [...] Insufficiency of recall and precision. [...] Heavy averaging. [...] Just documental retrieval” (INGWERSEN; JÄVERLIN, 2005, p. 7–9).

<sup>9</sup> “The central problems of information retrieval arise from the nature of the messages stored as records in the system and the relation of these messages to the queries likely to be put to the system” (VICKERY; VICKERY, 2004, p. 118).

<sup>10</sup> “In information retieval systems the “values” (for example, the texts) stored are of unlimited variety, the search terms presented in queries are unpretictable, and the relationship between messages stored and queries processed are often ambiguous” (VICKERY; VICKERY, 2004, p. 118).

Ferneda (2003) complementou essas afirmações, indicando que a maior dificuldade enfrentada pelos usuários é saber quais termos utilizar para encontrar documentos que atendam às suas necessidades específicas.

É importante lembrar que a recuperação de informação pode ser um processo subjetivo, uma vez que diferentes usuários podem ter diferentes maneiras de descrever a mesma necessidade de informação. Assim, Ferneda (2003) destacou que é fundamental que o SI permita uma certa flexibilidade na formulação de consultas, e ofereça recursos que possam ajudar os usuários a refinar suas buscas, como sugestões de termos relacionados ou filtros de resultados.

Em resumo, a recuperação da informação em ambientes digitais apresenta diversos desafios, como a carência de adaptação às necessidades e demandas específicas do usuário, a falta de interação e requisições dinâmicas, a insuficiência de variedade tática, a ausência de expectativa de incerteza, a carência de relevância orientada ao usuário, a falta de variedade dos bancos de dados, a questão da sobreposição documental, a insuficiência de recordação para precisão, o excesso de média nas buscas e a limitação à recuperação documental.

Além disso, a natureza das mensagens armazenadas no sistema e a relação dessas mensagens com as consultas dos usuários apresentam dificuldades, como a variedade ilimitada dos valores armazenados, a imprevisibilidade dos termos de pesquisa e a ambiguidade na relação entre as mensagens e as consultas. Dessa forma, o projeto de sistemas de recuperação da informação clássicos e digitais deve ter uma atenção depositada no usuário e sua relação com o sistema, a fim de aumentar as chances de vinculação entre busca e recuperação de informação.

Estas questões são agravadas e ampliadas quando se trata do projeto de sistemas de informação digitais. Isso ocorre porque a seleção dos documentos que devem ser armazenados no sistema é virtualmente inexistente, o que enfraquece a conexão entre as necessidades do usuário e o estoque de informações disponível. Além disso, quando o perfil e contexto de um usuário específico não são consideradas no projeto de sistemas de informação digitais há uma desconexão entre os registros do sistema e as necessidades efetivas dos usuários. Isso agrava a possibilidade de vinculação entre busca e recuperação da informação, comprometendo a relevância da informação recuperada. Além disso, o papel do usuário, como sujeito informacional em regimes de informação mais amplos, também tem um impacto na relação entre documentos e informação, conforme apontado por Rabello (2017).

De acordo com Gonzalez de Gómez (2012), a mesma combinação de tecnologia, serviços e conteúdos informacionais pode dar origem a múltiplos regimes de informação. Esses regimes estão relacionados às diferentes configurações que podem surgir entre os perfis individuais e sua interação com a informação. Para González de Gómez (2012), um regime de informação representa o modo predominante de produção de informação em uma sociedade, definindo os sujeitos envolvidos, as organizações, as regras, as autoridades informacionais, bem como os meios e recursos preferenciais de informação, padrões de excelência e modelos de organização, interação e distribuição.

Em ambientes que utilizam sistemas de informação digital, a introdução de um novo regime informacional pode impactar a definição dos sujeitos envolvidos, das instituições, das autoridades informacionais e dos recursos preferenciais de informação. Conforme explicado por Rabello (2017), esses regimes são mais flexíveis e adaptáveis, permitindo que o usuário desempenhe papéis tanto como fonte quanto como receptor de informação, tornando-se um sujeito informacional ativo.

Assim, nesses sistemas, os usuários também têm um papel de produtor, gerando dados e informações que alimentam os novos regimes informacionais (GONZALEZ DE GÓMEZ; RABELLO, 2017). Devido à quantidade, velocidade e variedade dessas produções, algoritmos de processamento são necessários e a dinâmica dos sistemas de acesso simples é alterada, tornando os usuários ativos na recepção de novas produções. Portanto, é de extrema importância examinar as maneiras de operação destes novos espaços, que são caracterizados por uma enorme produção de dados e usos de informações, agravando e intensificando problemas mencionados anteriormente e adicionando novas questões a serem consideradas.

#### **2.4 Aspectos gerais de sistemas de recuperação de informação digital**

Nos últimos anos, os ambientes informacionais digitais experimentaram uma transformação significativa na forma como manipulam e processam dados. Esta evolução pode ser parcialmente atribuída à incorporação de tecnologias inicialmente desenvolvidas para os ecossistemas chamados de *Big Data*. O conceito de *Big Data* é tradicionalmente definido pelas três Vs: Volume, Velocidade e Variedade

(DAVENPORT, 2014). Antes, estas eram considerações majoritariamente restritas a ambientes empresariais de grande escala e de pesquisa científica.

Graças ao desenvolvimento das TDICs, os usuários de sistemas de informação e recuperação de informação digitais não apenas consomem, mas também produzem dados, documentos e informações (RABELLO, 2017). Isso é possibilitado pelo uso de *smartphones*, sistemas de internet das coisas, redes sociais, *blogs* e outras ferramentas que alimentam estoques informacionais que são consultados por outros usuários. Esta nova dinâmica de relação entre o usuário e o SI resulta em um grande volume e velocidade de produção de dados e informações que, juntamente a utilização mais generalizada dessas ferramentas, pode levar a novos condicionantes nos fluxos informacionais, mais variados a nível de contexto de produção e uso.

Essa transição não é apenas uma mudança tecnológica, mas também conceitual. Enquanto as ferramentas clássicas de SIs e SRIs são, tradicionalmente, projetadas para operar sob um regime de previsibilidade e homogeneidade, a influência destas novas tecnologias introduziu uma nova complexidade. Agora, é imperativo que tais sistemas sejam capazes de operar eficientemente em cenários de alta imprevisibilidade, atendendo a uma gama cada vez mais diversificada de usuários com diferentes necessidades informacionais.

Para fazer isso de forma eficaz, tecnologias como aprendizado de máquina, análise em tempo real e armazenamento em nuvem foram adotadas para tornar tais sistemas mais robustos e ágeis. O resultado é uma redefinição dos paradigmas tradicionais de gestão de informação em ambientes digitais, em que as ferramentas oriundas do *Big Data* se tornam não apenas complementares, mas fundamentais para o funcionamento eficiente dos ambientes informacionais digitais modernos.

A incorporação de ferramentas derivadas do *Big Data* nos ambientes informacionais digitais, embora traga avanços, também impõe desafios substanciais, particularmente nos processos de recuperação de informação. Muitas dessas ferramentas operam com base em modelos estatísticos e algoritmos que foram projetados sob a suposição de uma certa homogeneidade nas bases de dados. No entanto, esta suposição frequentemente entra em conflito com a realidade dos ambientes informacionais contemporâneos. Nestes ambientes, a produção e o uso da informação são cada vez mais caracterizados por sua heterogeneidade e complexidade, ocorrendo em múltiplos contextos e atendendo a uma ampla gama de necessidades de usuário.

Tal dissonância entre a expectativa de homogeneidade dos modelos estatísticos e a heterogeneidade intrínseca da produção e consumo de informação introduz uma série de complicações na eficácia dos SRIs. Os sistemas são confrontados com o desafio de adaptar-se a dinâmicas imprevisíveis, exigindo maior flexibilidade e capacidades de personalização. Enquanto os modelos estatísticos buscam padrões e tendências gerais, a multiplicidade de contextos nos quais a informação é gerada e utilizada torna difícil para os SRIs oferecer resultados que são ao mesmo tempo precisos e relevantes para o usuário.

Desta forma, o estudo dos ambientes informacionais digitais é imperativo para compreender seus fluxos informacionais inerentes, especialmente em um cenário onde tais ambientes adotam largamente ferramentas antes associadas ao *Big Data*. Esta compreensão é importante para a adequação e otimização de SRIs, que frequentemente são desafiados pela heterogeneidade e a imprevisibilidade desses fluxos.

Em uma definição inicial, sistemas informação digitais são caracterizados pela recepção de uma grande produção de dados e documentos, concebidos em alta velocidade nos mais diversos contextos. De Mauro, Greco e Grimaldi (2016, p. 131, tradução nossa) qualificam este tipo ativo de informacional como

[...] caracterizado pelo alto volume, velocidade e variedade de produção, que demanda de tecnologias e métodos analíticos específicos para sua transformação em valor (DE MAURO; GRECO; GRIMALDI, 2016, p. 131, tradução nossa).<sup>11</sup>

Embora os SIs digitais tenham ganhado popularidade devido à sua aplicabilidade em uma ampla gama de setores e contextos, esta versatilidade pode ser paradoxalmente uma limitação. A estrutura técnica de tais sistemas, que se destina a capturar, armazenar e posteriormente tratar grandes volumes de dados de diversas fontes, pode falhar em fornecer informações direcionadas a perfis ou grupos específicos de usuários. Isso ocorre porque esses sistemas, ao contrário dos SIs e SRIs tradicionais, não possuem um mecanismo de classificação e indexação a priori, conforme apontado por Lima e Campos (2022) e Ferneda (2013), por exemplo. Assim,

---

<sup>11</sup> “*Big Data is the Information asset characterised by such a High Volume, Velocity and Variety to require specific Technology and Analytical Methods for its transformation into Value*” (DE MAURO; GRECO; GRIMALDI, 2016, p. 131).

mesmo que seu vasto repositório de dados e informações armazenadas pareça oferecer um potencial para gerar informações mais valiosas, a realidade é que eles muitas vezes resultam em saídas que não são eficazes ou direcionadas, comprometendo a eficácia de um SRI, conforme apontado por Barlow (2013).

A ausência de identificação prévia do usuário em potencial em SIs digitais apresenta desafios significativos para as técnicas estatísticas empregadas. Tais técnicas muitas vezes operam com amostras pouco direcionadas e genéricas, buscando identificar pontos comuns entre usuários que, na realidade, podem não representar adequadamente seus contextos específicos de uso da informação. Tal generalização compromete a capacidade do sistema em fornecer informações que sejam verdadeiramente relevantes e contextualizadas.

Esta limitação técnica aponta para a urgência de repensar e adaptar estes ambientes informacionais digitais. O grande volume de dados, que à primeira vista pode ser visto como uma vantagem, torna-se um problema quando a informação é armazenada de forma indiscriminada para tratamento posterior. Este procedimento contradiz o princípio de fornecer informações que sejam precisas e relevantes para contextos de uso distintos, conforme apontou Lima e Campos (2022), Smit e Barreto (2002) e Ferneda (2013). Assim, a eficácia destes sistemas em atender às necessidades informacionais específicas de seus usuários fica comprometida.

Sobre isso, Barlow (2013, p. 12, tradução nossa) destacou que, no passado,

[...] as infraestruturas de dados eram projetadas em torno de um conjunto conhecido de perguntas. Hoje em dia, é muito mais difícil prever as perguntas que serão feitas. Esta incerteza torna quase impossível construir infraestruturas no estilo tradicional para lidar com *Big Data* (BARLOW, 2013, p. 12, tradução nossa).<sup>12</sup>

Como resultado, pode haver uma lacuna entre as necessidades informacionais dos usuários e as informações recuperadas por estes sistemas. Sobre isso, Rabello (2013b, p. 162) destacou que os estudos de usuário para projeto de sistemas de informação digitais

[...] têm o desafio de reconhecer a necessidade dos usuários para adequar o sistema da melhor forma ao que é desejável, sem desconsiderar as possibilidades e os limites para mudanças e

---

<sup>12</sup> “*data infrastructures were designed around a known set of questions. Today, it’s much harder to predict the questions that will be asked. That uncertainty makes it nearly impossible to build traditional-style infrastructures for handling big data*” (BARLOW, 2013, p. 12)

aperfeiçoamentos. Pressupõe-se que, se os usuários não são conhecidos, logo, o sistema não poderá ser projetado de maneira a satisfazer as necessidades de quem usa (RABELLO, 2013b, p. 162).

A insuficiência na recuperação de informações nos sistemas modernos de informação não apenas afeta o indivíduo, mas pode também ter implicações mais amplas, influenciando grupos inteiros e até mesmo perpetuando disfunções sociais em grande escala (ZUBOFF, 2021). Esta realidade amplia a necessidade de estratégias mais sofisticadas e cuidadosas para a recuperação de informação, como apontam tanto O'Neil (2021) quanto Zuboff (2021). A abordagem atual, que muitas vezes carece de contextualização e precisão, pede uma revisão metodológica e técnica, em prol de sistemas mais eficazes e socialmente responsáveis.

A compreensão das técnicas operacionais é essencial, dado que as estratégias atuais em sistemas de informação digital baseiam-se em métodos anteriormente associados ao *Big Data*. Essas particularidades podem ser o fator subjacente à deficiência desses sistemas em fornecer informações de relevância específica para um usuário em particular. Portanto, é imperativo entender como essas técnicas funcionam para avaliar seu impacto na recuperação de informações.

#### **2.4.1 Técnicas e métodos de processamento de dados em sistemas de informação digitais**

A partir de uma perspectiva de análise operacional, os processos envolvidos no tratamento de dados de SIs digitais podem ser divididos em tecnologias e métodos (DE MAURO; GRECO; GRIMALDI, 2016). As tecnologias estão relacionadas à produção e armazenamento de dados, enquanto os métodos se concentram na geração de informações a partir destes dados, visando recuperar as informações solicitadas (DE MAURO; GRECO; GRIMALDI, 2016). No entanto, é importante destacar que as tecnologias (no sentido de técnicas) também são utilizadas para fins de análise e recuperação de dados, mas, para cada tipo de análise, é necessário considerar métodos específicos. As tecnologia e métodos considerados atentam para o desenvolvimento de sistemas de informação digitais no geral.

As técnicas de armazenamento que se destacam são aquelas que abrangem tamanho, volume, velocidade e variedade de contextos de produção. No entanto, dada a multiplicidade de contextos nos quais os dados são gerados em sistemas de

informação digitais, torna-se imperativo organizar esses dados de forma eficaz. Embora técnicas tradicionais de armazenamento de dados, como as baseadas em *Structured Query Language* (SQL), sejam eficazes para dados gerados em contextos mais organizados e específicos, elas mostram-se inadequadas para a complexidade e diversidade dos dados modernos.

Essa inadequação das tecnologias de bancos de dados relacionais, tradicionalmente utilizadas desde a década de 1970, é especialmente aparente quando se lida com dados que não se encaixam em modelos tabelares predefinidos (WHITE, 2015). A solução, portanto, tem sido o desenvolvimento e a adoção de bancos de dados *Not Only Standard Query Language* (NoSQL), como MongoDB, Cassandra, Couchbase, Amazon DynamoDB e OrientDB, para citar alguns.

Criados no início do século XXI, os bancos de dados NoSQL são projetados para acomodar dados gerados em uma ampla variedade de contextos, incluindo sistemas móveis e *Internet of Things* (IoT). Estes sistemas têm a capacidade de armazenar documentos inteiros e associá-los através de metadados ou outras tecnologias de nuvem. Eles compartilham características como flexibilidade de esquema, escalabilidade horizontal e capacidade de gerenciar dados distribuídos geograficamente (WHITE, 2015).

Muitas vezes, bancos de dados NoSQL são empregados em conjunto com outras plataformas como o *Hadoop Distributed File System* (HDFS), um sistema de arquivos distribuídos de código aberto da Apache, para lidar com grandes volumes de dados em ambientes distribuídos. O HDFS permite o acesso rápido e eficiente a dados, tornando-o uma solução adequada para os desafios apresentados pela diversidade de contextos de produção e complexidade dos dados atuais (WHITE, 2015). O HDFS possibilita a divisão dos dados em *clusters*, os quais são conjuntos de servidores, com o objetivo de facilitar e agilizar o acesso aos dados, além de garantir sua integridade. A plataforma realiza a seleção dos dados armazenados no BD e os distribui entre os servidores do *cluster*. Essa abordagem permite uma distribuição e indexação mais eficiente dos dados, tornando sua identificação e recuperação mais fácil. Em contraste, a centralização de processamento em um único servidor pode resultar em problemas de gerenciamento.

A distribuição dos dados em vários servidores proporciona um aumento na velocidade de processamento e simplifica a recuperação dos dados. Além disso, tal abordagem alivia a carga física de um BD centralizado em um único servidor. O HDFS,

por sua vez, garante a segurança dos dados por meio da duplicação das partições, caso ocorra uma falha em algum servidor. No entanto, é importante ressaltar que a distribuição de dados em vários servidores pode limitar a escalabilidade física do sistema, uma vez que requer a alocação de novos servidores interconectados em rede (MARQUESONE, 2016; WHITE, 2015).

Para solucionar esta problemática, serviços como *Azure*, *AWS* e *HDFS Cloud* apresentam soluções em nuvem, com virtualização para armazenamento de dados em grande volume, oriundos de diversos locais (BARLOW, 2013). Estes serviços possuem um conjunto de servidores online com serviços baseados em nuvem e em máquinas virtuais, capazes de promover certo grau de escalabilidade automatizada e distribuição indexada, mitigando a problemática de servidores locais, como no caso do HDFS.

Tanto soluções baseadas em nuvem quanto abordagens físicas manifestam deficiências na presença de mecanismos de identificação robustos e na qualidade dos conjuntos de dados institucionalizados. A etapa de seleção prévia emerge como um critério essencial no aprimoramento de sistemas de recuperação da informação. Este requisito se justifica pela necessidade de uma análise minuciosa do perfil do usuário e do contexto de uso da informação, conforme apontaram estudos de Smit e Barreto (2002), Lima e Campos (2022) e Ferneda (2013).

De fato, isso parece ser um fator deletério neste tipo de sistema, tendo em vista o grande volume de dados produzidos em alta velocidade com grande variabilidade midiática. A solução encontrada: – da seleção generalizada e posterior tratamento –, é largamente utilizada neste tipo de sistema, entendida como uma boa prática a ser replicada (VIDAL, 2019).

Sobre os métodos de processamento de dados em larga escala, existem modelos de processamento de dados em grande escala desenvolvidos para trabalhar em conjunto com sistemas de armazenamento NoSQL, como o *MapReduce* e o *Spark*. Estes métodos coletam dados armazenados e os estruturam, reduzindo o tamanho e criando correlações para apresentá-los da forma mais valiosa possível para um usuário. No entanto, a apresentação destes dados geralmente é feita em forma de tabelas e estruturas estatísticas, que devem ser organizadas e apresentadas de forma legível para o usuário, como informação (WHITE, 2015).

Para isso, ferramentas de *Business Intelligence* (BI), como o *Power BI* da *Microsoft* e o *Cognos Analytics* da *IBM*, atuam como mediadores, fornecendo

interfaces amigáveis para apresentar dados organizados em informações ao usuário. Essas ferramentas geralmente compilam as informações em formatos gráficos, permitindo interação na solicitação de novas informações, condição importante em SRIs clássicos (CHOO, 2003). Embora sejam úteis para a apresentação de dados, estes métodos carecem de desenvolvimentos que possibilitem apresentações e recuperação de informações mais orientadas as demandas reais de um usuário, em linguagem natural ou falada.

Mesmo sistemas *web*, projetados com base em boas práticas de organização informacional, como orientação a uma audiência específica ou a um conjunto de tarefas (MORVILLE; ROSENFELD, 2006), há dificuldade em apresentar a informação da melhor maneira possível, tendo em vista a variabilidade de usuários e demandas. Dessa forma, métodos capazes de lidar com este conjunto grande de dados foram e estão sendo desenvolvidos, alicerçados fortemente nas tecnologias de *ML*.

*ML* é um termo que pode ser confundido com Inteligência Artificial, mas, na verdade, é apenas um dos campos de estudo dentro dessa área mais ampla da TI (ARTERO, 2012; COPPIN, 2010). Em resumo, *ML* se refere a um conjunto de algoritmos que buscam identificar padrões em sistemas de dados e realizar previsões com base neles. Por outro lado, a Inteligência Artificial (IA) é um campo de estudo mais amplo, que se dedica ao estudo das faculdades mentais por meio de modelos computacionais ou ao estudo de projeto de agentes inteligentes (CHARNIAK; MCDERMOTT, 1985 *apud* NORVIG; RUSSELL, 2004, p. 5).

Diferentes métodos de *ML* podem ser empregados para processar dados e convertê-los em informações valiosas, especialmente, quando combinados com tecnologias como o *Spark* (WHITE, 2015). Em geral, o uso de técnicas de *ML* visa identificar padrões entre buscas e recuperação da informação, a fim de determinar quais padrões de busca estão mais relacionados aos padrões de recuperação da informação. A identificação destes padrões pode ser alcançada por meio de treinamento de ferramentas de aprendizado de máquina, como árvores de decisão, *k-means*, ou redes neurais artificiais (BITTENCOURT, 2006; BRAGA; CARVALHO; LUDERMIR, 2000; COPPIN, 2010). Essas ferramentas podem ser observadas em pacotes de linguagens de programação e processamento de dados conhecidas, como, por exemplo, o *TensorFlow*, *Keras*, *PyTorch*, *Scikit-learn*, *Apache Spark MLlib* e *Microsoft Azure Machine Learning* (GÉRON, 2019).

O *TensorFlow* é uma biblioteca popular de aprendizado de máquina e inteligência artificial. Ele é usado para criar e treinar modelos de forma eficiente. O *TensorFlow* funciona com base em um "grafo de fluxo de dados", onde as operações matemáticas são representadas como nós e os dados como tensores, que são basicamente matrizes multidimensionais. Você constrói um grafo computacional que define as operações e depois executa este grafo em uma sessão. O *TensorFlow* também pode ser distribuído em várias unidades de processamento para acelerar o treinamento e a execução dos modelos. É uma ferramenta poderosa com muitos recursos para ajudar no desenvolvimento de projetos de aprendizado de máquina (LOCASCIO; BUDUMA, 2017).

*Keras* é uma API de alto nível construída com base no *TensorFlow*, que permite criar e treinar modelos de aprendizado de máquina de forma mais simples e intuitiva. É uma ferramenta popular entre iniciantes em ML devido à sua facilidade de uso e capacidade de construir modelos complexos com poucas linhas de código (GÉRON, 2019).

O *PyTorch* é uma biblioteca de código aberto muito popular para aprendizado de máquina e inteligência artificial. O *PyTorch* foi desenvolvido para facilitar a criação e treinamento de modelos de forma eficiente. Uma das principais características do *PyTorch* é o seu foco em computação numérica utilizando tensores, que são basicamente matrizes multidimensionais. Esta biblioteca permite que você realize operações matemáticas complexas de maneira fácil e eficiente (HOWARD; GUGGER, 2020).

*Scikit-learn* é uma biblioteca de aprendizado de máquina de código aberto para *Python*, que oferece uma ampla variedade de algoritmos de aprendizado de máquina e ferramentas para pré-processamento de dados, seleção de recursos e avaliação de modelos. É uma ferramenta para tarefas de classificação, regressão e agrupamento (GÉRON, 2019).

O *Apache Spark MLlib* é uma biblioteca de aprendizado de máquina escalável e de código aberto. usado para realizar tarefas de análise de dados e construir modelos de aprendizado de máquina. Uma das principais características do *MLlib* é a sua capacidade de processar grandes volumes de dados de forma eficiente, aproveitando a tecnologia de processamento distribuído do *Apache Spark*. Isso permite lidar com conjuntos de dados muito grandes e realizar cálculos complexos de maneira rápida (ABBASI, 2017).

*Microsoft Azure Machine Learning* é uma plataforma de nuvem para desenvolvimento e implantação de modelos de aprendizado de máquina, oferecendo ferramentas para treinamento, avaliação e gerenciamento de modelos (GÉRON, 2019). Em cada uma dessas ferramentas, o treinamento é reforçado pelo usuário, que aprimora as relações, selecionando conjuntos de informações recuperadas para as palavras-chave previamente trabalhadas. Isso é feito de forma coletiva, ou seja, considera as combinações escolhidas por diversos usuários, em diversas iterações de busca e RI.

A combinação de ferramentas e métodos em sistemas de informação digitais tem como objetivo identificar padrões de repetição nas buscas e recuperações de informação, reforçando conjuntos de combinações para grandes grupos heterogêneos. No entanto, a falta de identificação de atributos específicos de usuários, de instrumentos de vinculação entre as informações armazenadas e perfis específicos, faz com que uma parcela significativa de dados seja desconsiderada na composição da informação. Isso ocorre, pois, concepções sobre como os sistemas inteligentes podem auxiliar nestes processos nem sempre estão alinhadas com as capacidades dos sistemas de IA emergentes. Isso demanda uma análise de como essas tecnologias emergentes devem ser efetivamente aplicadas, considerando fatores que vão além dos inicialmente considerados pela área de TI (JARRAHI *et al.*, 2023).

Além disso, os treinamentos em ferramentas de aprendizado de máquina em SIs digitais deixam de fora perfis de usuários que não fazem parte de grupos de treinamento, o que pode prejudicar a RI mais alinhada com perfis de usuários específicos, e deixar de lado informações relevantes para o usuário.

Uma solução para estes problemas é realizar uma avaliação contínua dos processos de recuperação da informação e aprimorar constantemente suas técnicas e métodos. Essa avaliação deve verificar o potencial de recuperação de informações relevantes para o usuário e avaliar se o sistema é capaz de realizar essa atividade com qualidade.

Nesta pesquisa, é proposta a utilização de atributos relacionados ao contexto de uso e perfil de usuário para aprimorar estes sistemas de informação digitais. Assim, é fundamental utilizar técnicas de avaliação de relevância das informações recuperadas, que serão abordadas em uma seção posterior juntamente ao conceito de relevância.

## 2.5 Relevância e pertinência da informação e avaliação de sistemas de recuperação da informação

Ao compreender e interpretar a informação recuperada, o usuário avalia seu potencial de relevância ou pertinência. Lima e Campos (2022) definiram relevância como o

[...] grau de similaridade entre termos que compõem as expressões de buscas de usuários e a ocorrência em documentos da coleção ou nos termos de indexação. Já a pertinência é a relação que existe entre a informação obtida em uma busca que responde à necessidade ou demanda de informação do usuário, ou seja, a informação que é útil para o usuário (LIMA; CAMPOS, 2022, p. 2).

Relevância, porém, é um conceito bastante subjetivo e elástico, dependendo do contexto em que é avaliado. A avaliação de relevância da informação depende de uma série de fatores objetivos e subjetivos dos usuários, que devem ser considerados para o projeto e aprimoramento de SIs. Sobre isso, Araújo Júnior (1995, p. 31) disse que este conceito é

[...] indissociável dos sistemas de recuperação da informação, do usuário e dos documentos, é extremamente subjetivo e sujeito a variações de interpretação e julgamento, dependendo dos momentos e condições iniciais do sistema, dos diferentes usuários e dos documentos em suas inter-relações. Qualquer alteração nessas variáveis pode mudar drasticamente os resultados esperados. O que é relevante para um elemento do sistema (responsável, por exemplo, pela seleção, indexação e demais processos) pode ou não ser para ele próprio em um outro momento no tempo, ou para outros elementos do sistema; o que é relevante para um usuário em um tempo T pode não ser para outros usuários ou para o mesmo em outro tempo T1; documentos têm sua própria relevância, a priori do sistema, e que pode ser alterada dependendo do conjunto ao qual esses documentos venham a pertencer; há, no sistema, uma imprevisibilidade de comportamento. E o sistema como um todo é sensível às alterações que tais imprevisibilidades vão provocar em suas variáveis – alterações estas que, conforme visto, não são lineares (ARAÚJO JÚNIOR, 1995, p. 31).

De forma geral, a relevância da informação pode ser considerada como um atributo de qualidade, ou seja, um atributo que indica se a recuperação de uma informação, em um SI ou SRI, atende as necessidades e as expectativas do usuário. Isso inclui aspectos como precisão, completude, atualidade, relevância, confiabilidade

e clareza da informação. Uma informação de alta qualidade é aquela considerada precisa, confiável, relevante e útil para o usuário em um contexto específico.

A percepção de relevância envolve diversos fatores além do atributo de qualidade, uma vez que os processos emocionais e lógicos envolvidos na percepção de relevância variam de acordo com as condições potenciais de uso da informação (KUHLTHAU, 1991). Identificar o que é relevante para um usuário específico é, portanto, condicionado por vários fatores complexos e inter-relacionados, tanto sistêmicos quanto relacionados à apropriação do uso da informação (ARAÚJO JÚNIOR, 1995).

Segundo Borlund (2003), duas abordagens podem ser consideradas para o estudo da relevância em SRIs: uma abordagem orientada pelo sistema, e outra abordagem cognitiva orientada pelo usuário, cada uma com sua própria compreensão de relevância.

A abordagem orientada pelo sistema trata a relevância como um conceito estático e objetivo, enquanto a abordagem cognitiva orientada pelo usuário considera a relevância como uma experiência mental subjetiva, individualizada, que envolve reestruturação cognitiva. Diferentes tipos de relevância são identificados em ambas as abordagens, incluindo relevância do sistema ou algorítmica, relevância semelhante ao tópico, pertinência ou relevância cognitiva, relevância situacional e motivacional, e afetiva (BORLUND, 2003).

A relevância algorítmica é o tipo mais comum e claro de relevância de sistema, é usada na avaliação tradicional de sistemas de RI, medindo quão bem o tópico da informação recuperada corresponde ao tópico da solicitação. Este tipo de relevância é restrito, pois lida apenas com o grau em que a representação da consulta corresponde ao conteúdo dos objetos de informação recuperado. A especificação dos tipos objetivos de relevância algorítmica pode ser rotulada como “tópico igual a conteúdo” (BORLUND, 2003).

As abordagens cognitivas, por sua vez, são mais direcionadas às particularidades dos usuários e, portanto, têm maior adequação para avaliar SRIs que lidam com perfis distintos de indivíduos. A abordagem cognitiva mais comum é a relevância situacional, que leva em conta a relação individual do usuário com a informação que foi recuperada. Diferentemente de outras abordagens, a relevância situacional não se baseia apenas na relação entre uma representação de consulta e um objeto de informação recuperado. Em vez disso, essa abordagem considera a

utilidade ou o valor do objeto de informação em relação à tarefa de trabalho em questão. A relevância situacional envolve aspectos motivacionais e afetivos, considera a característica de todos os tipos de relevância subjetiva, descrevendo a relação entre as intenções, objetivos e motivações do usuário, e os objetos de informação (BORLUND, 2003).

Embora a relevância situacional seja o conceito mais comumente considerado no estudo das relevâncias cognitivas, é importante mencionar outras abordagens complementares, que contribuem para a compreensão da subjetividade do usuário na avaliação de informações recuperadas. Na literatura, encontram-se os conceitos de "relevância psicológica" (HARTER, 1992), "relevância ostensiva" (CAMPBELL; VAN RIJSBERGEN, 1996) e "relevância da tarefa" (MIZZARO, 1998; REID, 1999).

A relevância psicológica, proposta por Harter (1992), está alinhada com as ideias básicas e fundamentais do ponto de vista cognitivo: a mudança das estruturas de conhecimento do receptor pelo ato de processamento de informações. A relevância psicológica descreve um estado de efeito que existe quando o usuário recupera informações, que sugerem novas conexões cognitivas, analogias frutíferas, metáforas esclarecedoras, aumento ou diminuição na força de uma crença. Assim, a relevância psicológica é vista como o efeito de uma mudança nas estruturas de conhecimento.

Em complemento, Campbell e van Rijsbergen (1996) propuseram o conceito de relevância ostensiva, que se refere ao grau em que as evidências do objeto de informação recuperado são representativas da necessidade atual de informação do usuário. Este conceito leva em conta a ideia de que a necessidade de informação é dinâmica, refletida na ponderação de probabilidade no "modelo ostensivo" de RI (CAMPBELL; VAN RIJSBERGEN, 1996). Tal modelo define a relevância como uma relação entre a consulta e o documento, em que um documento é considerado relevante se for útil para atender a necessidade de informação do usuário.

A relevância, segundo o modelo ostensivo, é uma característica do documento em relação à consulta, e não uma propriedade inerente do documento em si. Este modelo reconhece a subjetividade do usuário na avaliação da relevância e na escolha dos documentos, destacando a importância da interpretação da consulta pelo sistema, levando em consideração o contexto em que a consulta foi formulada e as necessidades do usuário (CAMPBELL; VAN RIJSBERGEN, 1996). Assim, o modelo ostensivo de recuperação da informação é importante para a compreensão da

subjetividade do usuário na avaliação da relevância e na escolha dos documentos mais adequados para sua necessidade de informação.

Em uma outra conceituação, Mizzaro (1998) propôs um modelo de relevância definida como a relação entre um recurso de informação e a representação do problema do usuário, avaliada de acordo com tópico, tarefa e/ou contexto em um determinado momento. Para Mizzaro (1998), o tipo final de relevância é a das informações recebidas para a necessidade real de informação do usuário em um determinado momento. Isso direciona a avaliação para o contexto de uso efetivo da informação, em detrimento do conceito de pertinência sugerido por Foskett (1972).

Reid (1999) complementou essa abordagem, chamando este tipo de relevância de “relevância da tarefa”, embora, sua definição seja idêntica à relevância situacional. Tanto Mizzaro (1998) quanto Reid (1999) estão interessados em capturar a utilidade percebida pelo usuário, os objetos de informação recuperados com referência à tarefa e necessidade real de informação

Schamber *et. al* (1990, p. 774, tradução nossa), de forma sumarizada, destacaram que a relevância da informação pode ser classificada em três categorias:

- a) Relevância é um conceito cognitivo multidimensional cujo significado depende, em grande parte, das percepções de informação dos usuários e de suas próprias situações de necessidade de informação.
- b) Relevância é um conceito dinâmico, que depende dos julgamentos dos usuários sobre a qualidade da relação entre a informação e a necessidade de informação em um determinado momento.
- c) Relevância é um conceito complexo, mas sistemático e mensurável se abordado conceitualmente e operacionalmente a partir da perspectiva do usuário (SCHAMBER, 1009, p. 774, tradução nossa).<sup>13</sup>

Ao considerar as abordagens mencionadas sobre o que pode ser considerado relevante para um usuário de SRI, é perceptível a presença de possíveis desafios para avaliar a informação recuperada, devido à subjetividade envolvida na avaliação, que

---

<sup>13</sup> “• *Relevance is a multidimensional cognitive concept whose meaning is largely dependent on users’ perceptions of information and their own information need situations.* • *Relevance is a dynamic concept that depends on users’ judgments of the quality of the relationship between information and information need at a certain point in time.* • *Relevance is a complex but systematic and measurable concept if approached conceptually and operationally from the user’s perspective*” (SCHAMBER; EISENBERG; NILAN, 1990, p. 774).

inclui aspectos emocionais, cognitivos e dinâmicos da relação entre o usuário e o sistema.

Assim, a avaliação de relevância algorítmica não é adequada para fornecer resultados satisfatórios neste contexto, tornando-se uma ferramenta inadequada para avaliar SIs digitais. De forma resumida, quatro pontos são percebidos como desafios para uma plena avaliação da relevância em sistemas de recuperação da informação: a avaliação da relevância, a avaliação de satisfação, a atualização de conhecimento de um usuário e a incerteza nas buscas por informação.

A avaliação da relevância de uma informação recuperada, por parte de um usuário, depende de certo conhecimento anterior sobre a temática ou a problemática de busca. Sem isso, o usuário pode ter dificuldade em encontrar valor ou relevância na informação recuperada. Sobre isso, Hjørland (2010, p. 231) apontou que

Determinar quais itens são relevantes em relação a um determinado objetivo/tarefa requer conhecimento do assunto e depende de diferentes teorias/visões. Os usuários de sistemas de informação, portanto, não são automaticamente competentes para julgar a relevância (HJØRLAND, 2010, p. 231, tradução nossa).<sup>14</sup>

Isso significa que a importância tanto da seleção quanto da produção de informações por um usuário depende do conhecimento prévio que ele tem sobre a informação em questão. No entanto, em contextos de produção de ambientes digitais com grande volume de dados, a análise de relevância se torna menos importante. Isso ocorre porque toda a produção é, geralmente, armazenada, dada a dinâmica das ferramentas digitais, que permitem que usuários comuns produzam dados e informações.

Durante a avaliação de satisfação do processo de recuperação de informações, é importante notar que nem sempre está relacionada à relevância da informação recuperada. Outros aspectos do processo, como desempenho do SI, apresentação da informação etc., podem influenciar na satisfação geral do usuário. Além disso, pode ser difícil, para o usuário, distinguir entre conceitos, como satisfação e relevância, e ainda, avaliar a qualidade da informação recuperada. Por fim, é importante ressaltar

---

<sup>14</sup> *“To determine which items are relevant in relation to a given goal/task requires subject knowledge and is dependent on different theories/views. Users of information systems are therefore not automatically competent to judge relevance” (HJØRLAND, 2010, p. 231).*

que a avaliação da relevância da informação recuperada nem sempre é o único fator que influencia na escolha do usuário, sobre quais registros utilizar.

Coeira e Vickland (2008) identificaram que, dependendo do processo de recuperação da informação, a avaliação da relevância pelos usuários não é um preditor forte do impacto das informações na tomada de decisão. Coeira e Vickland (2008) observam que a relação entre a relevância percebida e o impacto na tomada de decisão é complexa e deve ser investigada mais a fundo. Eles sugerem que outros fatores, como a credibilidade da fonte, a experiência do usuário e a confiança na tecnologia, também podem desempenhar um papel importante na tomada de decisões, conforme apresentado por Hjørland (2010) e Kuhlthau (1991). Schamber, Eisenberg e Nilan (1990, p. 760, tradução nossa) reforçam essa ideia:

[...] quando a satisfação é operacionalizada como uma medida na avaliação do desempenho de sistemas de informação, pode na verdade ser uma medida composta que contém vários tipos de julgamentos, incluindo julgamentos de relevância. A relevância e outros julgamentos (às vezes, o termo relevância nem é utilizado) podem ser usados para avaliar aspectos amplamente variados de um sistema (SCHAMBER; EISENBERG; NILAN, 1990, p. 760, tradução nossa).<sup>15</sup>

Também, a atualização do conhecimento de um usuário, tendo em vista a dinamicidade informacional, que pode alterar sua posição de relevância sobre uma mesma busca. Usuários com conhecimentos atualizados podem não mais ver relevância em informações as quais já possuem conhecimento. Hjørland (2010, p. 222, tradução nossa) afirmou que como

[...] o conhecimento está sempre atualizado, o próprio conhecimento muda dinamicamente e, portanto, a natureza dinâmica das “necessidades de informação” e da “relevância” é, em grande parte, causada por essa mudança em nosso conhecimento coletivo. Na literatura da Ciência da Informação, a natureza dinâmica da “relevância” está, no entanto, frequentemente ligada ao usuário, e não ao conhecimento em si (HJØRLAND, 2010, p. 222, tradução nossa).<sup>16</sup>

---

<sup>15</sup> “when satisfaction is operationalized as a measure in evaluating the performance of information systems, it may actually be a composite measure that contains several kinds of judgments, including judgments of relevance. Relevance and other judgments (sometimes the term relevance is not used at all) may be used to evaluate widely varied aspects of a system” (SCHAMBER; EISENBERG; NILAN, 1990, p. 760).

<sup>16</sup> “knowledge is always updated, knowledge itself changes dynamically, and therefore the dynamic nature of “information needs” and “relevance” is to a very large degree caused by this change in our collective knowledge. In the literature of information science the dynamic nature

Os aspectos afetivos, como confusão e incerteza na busca, podem também afetar o julgamento sobre a relevância de uma informação recuperada (KUHLETHAU, 1991). Em muitos casos, como visto, devido ao estado anômalo de conhecimento do usuário (COEIRA; VICKLAND, 2008), a avaliação da qualidade da informação recuperada, em termos de sua relevância, só pode ser feita após a apropriação e uso da informação. Além disso, barreiras linguísticas, cognitivas e simbólicas podem fazer com que uma informação seja descartada, devido à incapacidade do usuário de avaliar sua relevância. Isso se torna ainda mais desafiador, uma vez que sistemas de informação digitais, geralmente não são projetados para considerar os problemas dos usuários (BARLOW, 2013).

A fim de mitigar essas problemáticas, métodos de avaliação de relevância e pertinência da informação foram propostos em pesquisas como a de Araújo Júnior (2005), Manning, Raghavan e Schütze (2008) e Fawcett (2005). Estes métodos visam à avaliação da resposta do usuário frente a uma informação recuperada, fornecendo *feedback* para o sistema, com objetivo de calibrá-lo na relação de busca e recuperação da informação.

Uma das maneiras de realizar tal avaliação é por meio de atributos de índice de precisão e revocação (MANNING; RAGHAVAN; SCHÜTZE, 2008). Segundo Araújo Júnior (2005), precisão é

[...] um conceito fundamental para a avaliação da qualidade da recuperação da informação, ao mesmo tempo que representa a medida de interesse (informação útil) do que foi encontrado em um processo de busca e recuperação da informação para o usuário, que qualifica a informação recuperada como útil ou inútil de acordo com as suas necessidades (ARAÚJO JÚNIOR, 2005, p. 91).

De acordo com Lancaster (1998), a precisão se refere à medida que os itens recuperados em um processo de busca e recuperação de informações, em uma base de dados, são considerados relevantes pelo usuário. A alta precisão é alcançada quando a maioria ou todos os itens bibliográficos recuperados são considerados pertinentes pelo usuário (ARAÚJO JÚNIOR, 2005).

---

*of "relevance" is, however, often connected to the user, rather than to knowledge itself* (HJØRLAND, 2010, p. 222).

Araújo Júnior (2005, p. 92) exemplificou a questão da precisão indicando que esta

[...] pode ser determinada por uma proporção, conclusão a que se chega em pesquisas feitas na literatura corrente: se 50 documentos são recuperados e o usuário julgar que 10 são úteis, o índice de precisão será calculado pela proporção de 10/50, que dará um resultado de 20% (ARAÚJO JÚNIOR, 2005, p. 92).

A precisão de um processo de busca e recuperação de informação, em uma base de dados, é diretamente influenciada pela avaliação do usuário em relação aos itens recuperados. Se todos os itens recuperados forem considerados úteis pelo usuário, a precisão será de 100% (ARAÚJO JÚNIOR, 2005). Porém, se uma parcela significativa dos itens não for considerada útil, a precisão será menor. Em resumo, a precisão é uma medida da qualidade da recuperação de informação e está diretamente relacionada à relevância dos itens recuperados para o usuário.

Tradicionalmente, a precisão é vista como inversamente proporcional à revocação, o que significa que, quanto maior for a precisão, menor será a revocação (ARAÚJO JÚNIOR, 2005).

A revocação “pode ser conceituada como a porcentagem de itens úteis que um termo ou combinação de termos, pode recuperar” (ARAÚJO JÚNIOR, 2005, p. 93). No contexto de uma busca realizada por um usuário, em uma coleção de documentos,

[...] a revocação será dada se houver acesso ao documento necessário, logo após o processo de busca e recuperação da informação. Se há uma busca por termos em uma base de dados, a revocação é medida de acordo com o número de referências úteis recuperadas (ARAÚJO JÚNIOR, 2005, p. 93).

Araújo Júnior (2005), com base em Lancaster (1998), apresentou uma relação matemática passível de ser utilizada para a avaliação dos processos de recuperação de informação em um SRI que pode ser observada na Figura 8.

**Figura 8 - Relação matemática entre os atributos de revocação e precisão em uma recuperação de informações**

$\text{Relação da revocação} = \frac{A \cap B}{A} = \frac{\text{(documentos úteis recuperados)}}{\text{(total de documentos úteis)}}$
$\text{Relação da precisão} = \frac{A \cap B}{B} = \frac{\text{(documentos úteis recuperados)}}{\text{(total de documentos recuperados)}}$

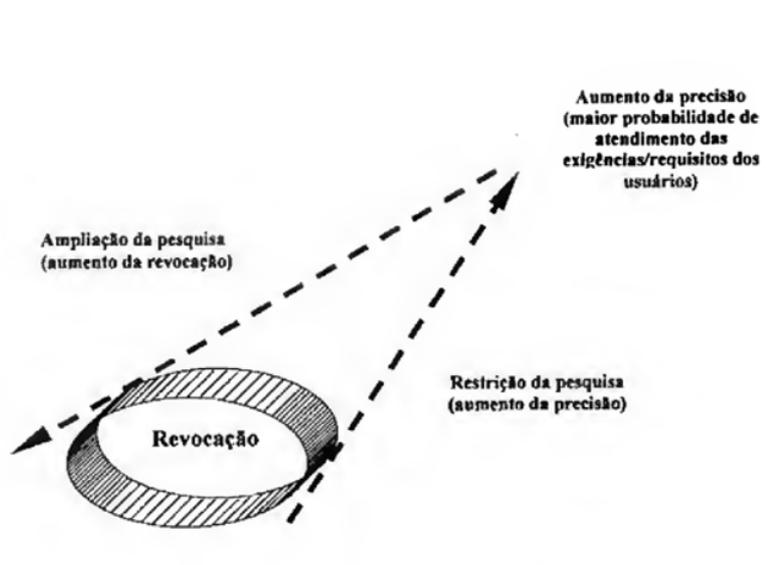
Fonte: Araújo Júnior (2005, p. 95).

Sendo “A”, o total de documentos úteis recuperados; e “B”, o total de documentos recuperados.

É importante observar que a relação entre revocação e precisão é inversamente proporcional, ou seja, quando a revocação aumenta, a precisão tende a diminuir, e vice-versa. Esta é uma das principais proposições na avaliação da recuperação da informação, indicando que a busca ideal deve ter alta precisão e alta revocação (ARAÚJO JÚNIOR, 2005).

A Figura 9, também de Araújo Júnior (2005, p. 96), explicita essa proporcionalidade.

**Figura 9 - Relação entre os atributos de revocação e precisão em uma recuperação de informações**



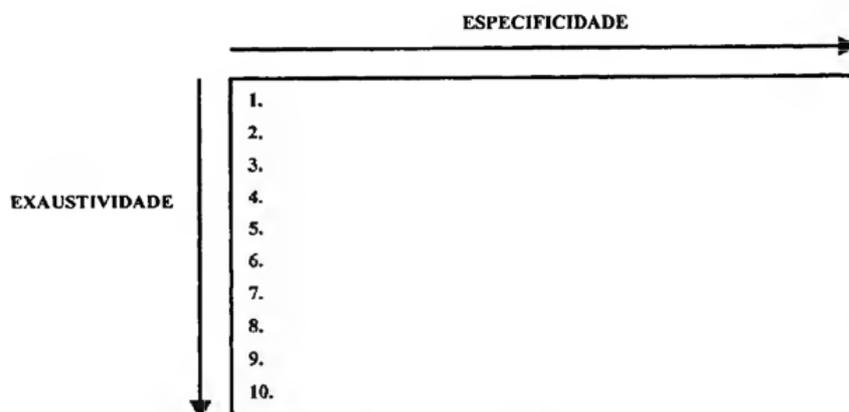
Fonte: Araújo Júnior (2005, p. 96).

Outros dois conceitos são fundamentais para a compreensão da precisão em sistemas de recuperação de informação: especificidade e exaustividade. A especificidade está atrelada à minuciosidade da indexação da informação, afetando diretamente a precisão das buscas. Quanto maior a especificidade, maior a chance de recuperar informações altamente relevantes (ARAÚJO JÚNIOR, 2005). Por outro lado, a exaustividade refere-se à amplitude com que um documento é analisado para determinar seu conteúdo temático durante o processo de indexação. Essa é uma decisão administrativa que define a extensão em que a informação é abordada, influenciando tanto a precisão quanto a revocação dos resultados da busca (MANNING; RAGHAVAN; SCHÜTZE, 2008). Ambos os conceitos, portanto, são cruciais para a eficácia de sistemas de recuperação de informação, impactando diretamente as métricas de desempenho desses sistemas. Sobre exaustividade, Araújo Júnior (2005) apontou que

[...] considerando o objetivo da abordagem da especificidade e da exaustividade na análise da precisão, em princípio fica claro que o aumento da exaustividade pode trazer pouca ou nenhuma vantagem no incremento da precisão, em oposição à especificidade que poderá trazer benefícios (ARAÚJO JÚNIOR, 2005, p. 96-97).

A Figura 10, também de Araújo Júnior (2005, p. 97) demonstra a relação entre exaustividade e especificidade.

**Figura 10 - Relação entre a especificidade e a exaustividade da informação recuperada**



Fonte: Araújo Júnior (2005, p. 97).

A relação entre exaustividade e especificidade em SIs é essencialmente uma troca. Aumentar a exaustividade, ou seja, incluir mais informações nos resultados da pesquisa, geralmente reduz a especificidade, tornando os resultados menos precisos e potencialmente menos relevantes para o usuário. Por outro lado, buscar maior especificidade limita o escopo dos resultados, tornando-os mais precisos, porém, potencialmente menos abrangentes. Tal dinâmica exige uma consideração cuidadosa das necessidades e objetivos do usuário ao configurar sistemas de informação para garantir que a relação entre exaustividade e especificidade seja adequada para a tarefa ou contexto específico.

Ao avaliar o processo de recuperação da informação, o usuário pode, por meio destas métricas, prover instrumentos para a melhoria, a longo prazo, do SRI, permitindo que este recupere informações mais relevantes e úteis. Sobre isso, Araújo Júnior (2005, p. 105) afirmou que, ao

[...] focar a satisfação do usuário como sendo o objetivo final do processo de gerenciamento da informação, o processo de busca e recuperação da informação deverá planejar a ação de incremento da precisão, a partir de tarefas a serem executadas durante todo o processo (ARAÚJO JÚNIOR, 2005, p. 105).

Tais métricas permitem uma análise mais precisa da eficiência do sistema em recuperar informações relevantes e de sua capacidade em cobrir toda a informação relevante disponível. Ao considerar o contexto de uso da informação e o perfil do usuário, é possível ajustar e aprimorar o design do SRI, proporcionando uma melhor experiência de busca e recuperação de informações para os usuários. Para identificar quais atributos contextuais são considerados melhores para o aprimoramento e posterior avaliação, realizou-se uma revisão de literatura sobre os diversos aspectos que constituem o conceito de contexto de uso da informação e características do usuário, apresentados na próxima seção.

## **2.6 A noção de contexto**

O processo de recuperação da informação é único para cada usuário, além se instância de busca, realizada em um SRI. Aspectos diversos podem influenciar na maneira como uma informação é buscada e utilizada por um usuário, considerando as especificidades de seu perfil e demanda. Entende-se, de forma inicial, que a

sobreposição e a combinação destes diversos aspectos, ou características, que impactam no momento de uso de um SRI, são considerados como o contexto de uso da informação (ABOWD; DEY, 1999). Sobre isso, Cool e Spink (2002, p. 605, tradução nossa) dizem que

[...] é geralmente reconhecido que a recuperação da informação é um processo intrinsecamente interativo, que ocorre dentro de múltiplos contextos sobrepostos que informam, direcionam ou moldam a natureza dessa interação. Em outras palavras, a busca, uso e avaliação da informação ocorrem dentro de contextos multidimensionais, que podem ser analisados em múltiplos níveis. (COOL; SPINK, 2002, p. 605, tradução nossa).<sup>17</sup>

Neste trabalho, considerou-se esses contextos multidimensionais para identificar o usuário da informação, bem como seu problema informacional, em um SI digital. Não se pode descartar a existência de um contexto de produção da informação. Buckland (1991) indicou que a informação é circunstancial, e os documentos portadores da informação são informativos, conforme as circunstâncias, tanto no nível de uso quanto no de produção.

Este trabalho teve como objetivo principal desenvolver um modelo que melhore as respostas de um SRI em ambientes digitais, incorporando elementos do contexto de uso. O foco esteve na otimização da eficácia e pertinência das respostas do SRI, reconhecendo os desafios enfrentados na recuperação de informações. Para atingir tal meta, a pesquisa explorou o contexto de uso da informação, abordando a situação do usuário durante a consulta, incluindo espaço e tempo, bem como o subsequente uso da informação, levando em consideração esses mesmos aspectos. Dessa forma, este estudo concentrou-se na compreensão do contexto de uso da informação e nas características dos usuários, considerando diversos contextos de uso da informação. Conforme explicitado por Foresti, Varkakis e Godoy Vieira (2016, p. 6):

---

<sup>17</sup> *'it is generally recognized that IR is an inherently interactive process, which occurs within multiple, overlapping, contexts that inform, direct or shape the nature of this interaction. In other words, information seeking, use and evaluation take place within multidimensional contexts, which can be analyzed from multiple levels'* (FORESTI; VARVAKIS; GODOY VIERA, 2016).

O contexto se mostra importante para compreender o usuário e as diversas sociedades e culturas, bem como os problemas enfrentados pelas unidades de informação (FORESTI; VARKAKIS; GODOY VIEIRA, 2016, p. 6).

Este contexto, normalmente, está alicerçado a condições sócio-temporais, que determinam o viés histórico e cultural da informação. Sobre isso, Marteleto, Nóbrega e Morado (2013, p. 100) declaram que o “conhecimento que resulta da ação e das relações dos sujeitos em seu fazer na história, em suas práticas culturais, é sempre um conhecimento contextual”. Capurro e Hjørland (2003) complementaram tal afirmação, indicando que, ao discutir o termo informação, é necessário ter em mente não apenas sua conotação de novidade ou relevância, mas também, reconhecer que a informação é um processo de transformação do conhecimento, envolve seleção e interpretação em contextos específicos.

Nota-se, assim, que os fluxos informacionais acontecem em variados contextos, indicando condicionantes específicos temporais, locais, culturais e afins da produção e uso da informação. Dumont e Gattoni (2003, p. 47) reforçam este posicionamento, dizendo que

[...] se os contextos sociais não forem levados em suficiente consideração, as pesquisas e estudos em Ciência da Informação tornam-se limitados, no mínimo distorcidos, no tocante às pesquisas do fenômeno informacional (DUMONT E GATTONI, 2003, p. 47).

É importante salientar que o contexto de uso da informação não é atributo necessariamente rígido, assim como pode auxiliar a qualificar um usuário, pode ser determinado por ele. Sobre isso, Araújo Júnior (2017, p. 229) afirmou que o

[..] usuário, assim, não é totalmente determinado pelo contexto no qual se insere, nem é totalmente isolado ou alheio a ele; a determinação que o contexto exerce existe, é real, mas não é mecânica nem absoluta, é interpretada e alterada pelo sujeito (ARAÚJO JÚNIOR, 2017, p. 229).

Isso faz com que o estudo dos contextos de uso da informação deva ser feito não somente à luz de um objetivo taxonômico, mas também, como condicionante qualificador do usuário em sua especificidade cultural, temporal, espacial.

Uma vez que SIs digitais, não possuem ferramentas técnicas capazes de avaliar plenamente as demandas dos usuários (BARLOW, 2013), é importante explorar o conceito de contexto. Isso pode ajudar a identificar, de maneira mais eficiente, as necessidades e preferências dos usuários, como já é feito no desenvolvimento de sistemas de informação clássicos. Henrique, Nassif e Venâncio (2007, p. 8) reforçaram tal afirmação, complementado que, para o projeto de sistemas de recuperação de informação,

[...] torna-se cada vez mais importante considerar as condutas dos sujeitos no domínio de ação, ou contexto (escola, trabalho, família), ou ainda, situação específica em que estão sendo observados; os contatos que estabelecem no cotidiano e como e porque eles ocorrem; as pré-disposições e interesses relacionados ao domínio de ação no qual estão sendo observados (assuntos relacionados, publicações específicas, etc.) e a pré-disposição emocional e afetiva, desses usuários nos domínios de ação que determinam o grau de interação sujeito-meio (HENRIQUE; NASSIF; VENÂNCIO, 2007, p. 8).

Dey (2000) indicou que, para que seja capaz de recuperar a informação mais apropriada para um usuário, um SI deveria possuir atributos de sensibilidade ao seu contexto, ou ser *context-aware*. Isso significaria que um SI deveria ser capaz de ser mais sensível a essas especificidades relativas ao uso da informação e seu usuário, recuperando uma informação mais alinhada à efetiva realidade de uso, para além das palavras-chave ou termos utilizados para buscar uma informação. Dessa forma, SIs ou SRIs devem ser projetados com capacidade de identificação de características contextuais, utilizando-as como atributo para refinamento do processo de RI, promovendo a percepção do estado em que se dá uma busca de informação.

SIs clássicos, em especial, aqueles mediados por atores humanos, realizam esta ação de captação e utilização de atributos contextuais através de um processo de negociação. Neste processo, há interação com o usuário, em que o mediador o indaga sobre sua necessidade de uso da informação, sensibilizando-se acerca das nuances que vão para além das impressões de necessidade relatadas. Tal sensibilização desenha o cenário em que a informação potencialmente será utilizada, de que forma e por quem, amplificando as chances de a informação recuperada ter relevância para o usuário, dentro de seu contexto. A percepção das nuances, que constituem os atributos contextuais no momento da busca da informação, é tarefa complexa, tendo em vista que, por vezes, demanda captação de características subjetivas.

Um intermediador humano é capaz de perceber, a partir de atributos de perfil de um usuário, melhores formas de apresentação da informação, bem como, quais informações podem ser mais apropriadas, dadas as especificidades de seu uso. Este conjunto de capacidades, formadas por tempo de experiência, por vezes, não é percebido ou acionado de forma explícita e consciente, diluindo-se durante o processo de mediação feita por um ator humano.

No contexto mencionado, a tentativa de captação e utilização dos fatores contextuais, mesmo que sejam subjetivos, são necessárias. Neste sentido, três problemas se destacam: (1) quais atributos contextuais devem ser considerados durante o processo de recuperação da informação a fim de aumentar o potencial de relevância de uma informação? Esse problema diz respeito à identificação dos elementos do contexto que são relevantes para determinar a relevância de uma informação. Exemplos de atributos contextuais podem incluir a localização geográfica do usuário, o horário da consulta, o histórico de interações anteriores, entre outros. (2) como estes atributos contextuais devem ser coletados? Neste caso, trata-se da definição das estratégias e técnicas para obter estes atributos do usuário ou do ambiente em que a consulta está ocorrendo. Isso pode envolver a solicitação direta de informações, a coleta automática de dados por meio de sensores ou a utilização de técnicas de mineração de dados para extrair informações contextuais a partir de outros recursos disponíveis. (3) como estes atributos contextuais devem ser traduzidos em algoritmo para aprimorar o processo de RI em SI digital? Aqui, o desafio é converter os atributos contextuais coletados em regras, modelos ou técnicas algorítmicas que possam ser incorporados no sistema de RI. Isso envolve o desenvolvimento de algoritmos que considerem estes atributos no processo de classificação, ranqueamento ou filtragem das informações, de modo a melhorar a precisão e a relevância dos resultados apresentados aos usuários.

A fim de responder estas questões, para dar base ao desenvolvimento do modelo proposto no trabalho, foram consultadas referências teóricas que versam sobre a utilização de contexto em SIs, bem como, sua transversalidade com sistemas digitais baseados em algoritmos computacionais.

Em uma primeira consulta, a norma NBR 9241-11 (ABNT, 2021), aborda questões relacionadas à usabilidade de computadores, fornecendo visões sobre problemas no projeto de sistemas relacionados à computação. A referida norma abrange uma ampla gama de considerações relacionadas à usabilidade, englobando

aspectos ergonômicos, interação humano-computador e design de interfaces. Seu enfoque principal é assegurar que os sistemas sejam concebidos de forma a atender plenamente às exigências e anseios dos usuários, promovendo assim uma interação com o computador que seja altamente eficaz, eficiente e satisfatória. Esses princípios da norma se mostram igualmente pertinentes e aplicáveis ao contexto de projetos de SIs, onde a usabilidade desempenha um papel importante na qualidade e no desempenho dos sistemas.

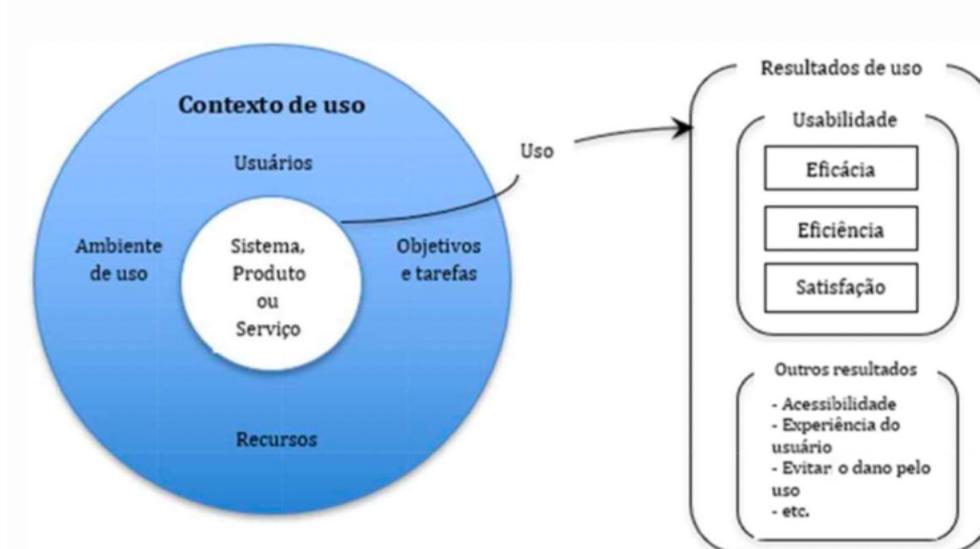
Sobre contexto, seu posicionamento se assemelha ao exposto anteriormente, pois entende que as peculiaridades do uso de um sistema por parte de um usuário afetam sua usabilidade. Importante frisar, no entanto, que essa é uma aproximação adotada, pois essa conceituação de contexto é passível de utilização na pesquisa. A norma trata de usabilidade, em especial àquela relacionada à projetos que demandem de orientações sobre ergonomia. Porém, sua adaptação para o projeto de SIs é possível, tendo em vista que também considera a usabilidade de sistemas. A norma, neste sentido, apontou que o

[...] contexto de uso consiste de usuários, tarefas, equipamentos (hardware, software e materiais), e do ambiente físico e social, pois todos esses podem influenciar a usabilidade de um produto dentro de um sistema de trabalho (ABNT, 2021, p. 2).

Cabe, no entanto, indicar que seu desenvolvimento está alicerçado em práticas da TI. Apesar disso, é possível considerar seu posicionamento sobre usabilidade como ponto inicial de análise. Como observado, a norma condiciona o projeto de sistemas computacionais a seus usuários, entendendo que estes possuem peculiaridades que afetam o desenvolvimento dos sistemas. Isso, *a priori*, alinha-se com o observado em Foresti, Varkakis e Godoy Vieira (2016), Marteleto, Nóbrega e Morado (2013) e Henrique, Nassif e Venâncio (2007), por exemplo.

Sobre usabilidade, relacionada com o contexto de uso do sistema, a norma indica a forma como um sistema computacional deveria operar, no sentido de atingir um objetivo específico. A Figura 11 apresenta a estrutura de usabilidade descrita na norma, evidenciando a preocupação com o contexto de uso do produto, no caso da presente pesquisa, a informação.

**Figura 11 - Estrutura da usabilidade de computadores, com indicação de contexto de uso, conforme norma NBR 9241-11**



Fonte: ABNT (2021, p. 8).

A norma enfatiza a importância da consideração do contexto de uso ao projetar um sistema computacional, destacando quatro pontos de atenção: usuário, tarefa, equipamento e ambiente.

Sobre usuários, a norma NBR 9241-11 (ABNT, 2021) indica que pode ser descrito a partir de seus conhecimentos, habilidades, experiência, educação, treinamento, atributos físicos e capacidades sensoriais e motoras. Indica, também, que há peculiaridades que devem ser consideradas na descrição de um usuário específico. Sobre a forma de captação e análise destes atributos, a norma apresenta um sistema de tabulação básico, não orientando sobre métodos de levantamento e análise. Este usuário, novamente, é o considerado para no projeto de sistemas e produtos que se relacionam com operações que demandem de orientação sobre ergonomia. É reforçado, no entanto, o potencial de aproximação da norma com o projeto de SIs, dada similaridade nos conceitos de contexto.

Sobre tarefas e equipamentos, a norma versa sobre o que se espera que um usuário faça com um sistema computacional, bem como, quais devem ser as características necessárias de *software* e *hardware* para que um sistema seja capaz de operacionalizar as tarefas descritas. Este levantamento deve levar em conta as possibilidades de ação previstas e não previstas de um usuário frente ao sistema.

A questão ambiental, presente na norma, destaca a importância do levantamento de atributos e análise de condições físicas e sociais, que cercam o produto (informação) e o usuário. As questões físicas estão relacionadas ao ambiente de uso, bem como, mobiliário, térmico, temporal etc. As questões sociais que devem ser levantadas têm a ver com práticas sociais, culturais, atitudes, entre outros. Há relação, novamente, do usuário com estes atributos, essencial para o entendimento do contexto em que um sistema computacional deve ser projetado.

A variação de combinação destes pontos, conforme apresentado na norma, explicita um posicionamento sobre o contexto de uso.

Cabe destacar, no entanto, que, ao debruçar sua atenção para o desenvolvimento de sistemas computacionais, a norma direciona o entendimento destes atributos como entidades para o desenvolvimento de projetos. Sistemas de informações digitais, em especial, aqueles que trabalham com grandes volumes de dados, não devem considerar as condições contextuais, conforme apontado por Araújo (2017), de forma isolada. Há interferência do usuário no que tange seu contexto em cada caso, e SIs devem ser tenazes em relação a essa condição.

Foresti, Varkakis e Godoy Vieira (2016, p. 8) reforçaram este posicionamento, dizendo que é importante que a CI compreenda os mecanismos que promovem a desarticulação contextual da informação e do conhecimento, bem como, o impacto deste fenômeno nos usuários e nas instituições, considerando que o “ contexto pode interferir na apropriação e uso da informação pelos usuários”. Assim, há necessidade de atenção do posicionamento da CI sobre o contexto, tendo em vista que o projeto de SIs, orientado ao atributo, deve ser dinâmico e flexível.

Em uma primeira perspectiva dentro da CI, Talja, Heidi e Tarja (1999) nos apresentam duas abordagens: a objetificada e a interpretativa. Na abordagem objetificada, podem ser observados esforços para analisar e caracterizar os possíveis fatores ou atributos que influenciam o processo de busca e recuperação da informação, de forma geral, sem relacioná-los. Nesta abordagem, as características contextuais representam o ambiente de busca e recuperação, sendo comuns para qualquer tipo de busca e recuperação.

Há, nesta abordagem, uma questão taxonômica, de tentativa de classificação geral e coletiva, sem levar em conta a realidade individual do usuário (TALJA; HEIDI; TARJA, 1999). Nela, os estudos sobre contexto apresentam a qualidade de parecerem “simples e transparentes”, sem grandes impedimentos de compreensão

dos fatores contextuais. Ainda, as pesquisas alicerçadas nessa abordagem adotam a ideia de “mais é melhor”, entendendo que quanto maior forem os volumes de dados e informações sobre o contexto em um determinado momento de busca, maior será a precisão no processo de RI de um usuário. Sobre isso, as autoras indicam que

[...] uma noção objetificada de contexto, muitas vezes, considera que quanto mais variáveis os fatores contextuais forem levados em consideração, maior é a certeza sobre a natureza do objeto de pesquisa.<sup>18</sup> (TALJA; HEIDI; TARJA, 1999, p. 155, tradução nossa).

A norma NBR 9241-11 (ABNT, 2021) se enquadra neste tipo de abordagem, ao entender que aspectos contextuais fazem parte de uma estrutura taxonômica, tabelável.

A relação entre usuário e sistema, no entanto, também deve ser levada em conta. Entende-se que essa relação demanda de um método mediador entre a necessidade do usuário e as possibilidades de ação do sistema. Neste caminho, a abordagem interpretativa se eleva.

Na abordagem interpretativa, os atributos ou as características contextuais identificadas referenciam a forma como a busca e a recuperação da informação seria mediada, variando para cada usuário e sistema. Isso quer dizer que, dadas as condições contextuais no momento de busca e RI, diferentes atributos contextuais poderiam ser observados (TALJA; HEIDI; TARJA, 1999). Isso se assemelha ao exposto anteriormente, em especial, ao apontado por Araújo (2017). As autoras dizem que, na

[...] interpretativa, os dados não são entendidos como uma descrição direta da realidade; em vez disso, os dados representam a realidade social. Nesta visão, observações, entrevistas, diários e materiais textuais, como artigos, representam diferentes contextos de interação e construção de significados, mas nenhum deles é uma descrição mais autêntica ou não mediada da realidade [...] (TALJA; HEIDI; TARJA, 1999, p. 755, tradução nossa).<sup>19</sup>

---

<sup>18</sup> *‘Within an objectified notion of context, it is often considered that the more variables or contextual factors that are taken into consideration, the higher the certainty about the nature of the research object’ (TALJA; HEIDI; TARJA, 1999, p. 155).*

<sup>19</sup> *‘interpretative approach, data are not understood as straightforward description of reality; rather, data represent social reality. In this view, observation, interviews, diaries and text materials like articles represent different contexts of interaction and sensemaking, but none are more of an authentic or unmediated description of reality [...]’ (TALJA; HEIDI; TARJA, 1999, p. 755).*

Dada a condição relativa dos sistemas de informação digitais de possuir características dinâmicas, compreender o contexto neste tipo de ambiente se torna importante. Em modelagem, a identificação de formas de captação e identificação contextual, mesmo de forma dinâmica, faz-se também necessária. Talja, Heidi e Tarja (1999) não medem esforços para tentar apresentar classificações possíveis de atributos para o projeto de sistemas de informação, focando nessas diferentes abordagens.

Assim, em complemento, Foresti, Varvakis e Godoy Viera (2016), apresentaram a visão linguística de contexto, que envolve também questões simbólicas e subjetivas, alinhando-se, em parte, à abordagem interpretativa, descrita por Talja, Heidi e Tarja (1999). Assim, há um esforço classificatório contextual para uso e análise dos mais diferentes tipos de SIs.

Foresti, Varvakis e Godoy Viera (2016) indicaram que os atributos contextuais podem ser classificados em três grandes grupos: (1) linguístico/verbal, que trata do conhecimento sobre uma forma de comunicação; (2) situacional, que trata da junção das estruturas de comunicação e seu aporte social de uso; e (3) extralinguístico, que lida com questões ligadas a crenças, valores e conhecimentos prévios.

Sobre o conhecimento das formas de comunicação, Foresti, Varvakis e Godoy Viera (2016) indicaram que estes meios podem ser representados, contextualmente, como

[...] “texto” onde as palavras ou frases (“unidades linguísticas”) estão inseridas. Esse texto pode ser oral ou escrito. Pode-se entender esse tipo de contexto como o suporte da informação, entender o “texto” como filmes, sites, gravações sonoras (FORESTI; VARVAKIS; GODOY VIERA, 2016, p. 4).

Sobre a questão situacional, Foresti, Varvakis e Godoy Viera (2016) indicaram que seu conhecimento pressupõe familiaridade com o sistema sógnico/simbólico em que a mensagem, dentro do fluxo informacional, foi construída. Isso quer dizer que os atributos contextuais relativos a este grupo dizem respeito à familiaridade do usuário com os instrumentos de comunicação. Sobre isso, estes atributos dizem

respeito à soma das “informações que o usuário da língua” tem acerca do uso das unidades linguísticas em seu “contexto social”. Essa definição mostra toda a complexidade do conceito, que envolve o conhecimento do usuário sobre a língua e a sua habilidade de usar esse conhecimento em suas relações sociais (FORESTI; VARVAKIS; GODOY VIERA, 2016, p. 4).

A questão extralinguística, conforme apontado por Foresti, Varvakis e Godoy Viera (2016), se posiciona abaixo da questão situacional, que trata do conjunto de conhecimentos prévios de um usuário sobre o que se informa, bem como, crenças, valores etc. Pressupõe-se que um usuário possui algum grau de conhecimento sobre aquilo que pesquisa, bem como, carrega sua pesquisa como parte de si, a partir de suas necessidades individuais. Neste sentido, com

[...] respeito aos “dados extralinguísticos” importantes a realização do “ato de comunicação linguística”. Envolve a “situação imediata” de aplicação de determinado “enunciado”, a soma dos “conhecimentos” do emissor e do receptor sobre o assunto. Também envolve as “crenças e pressuposições subjacentes” ao assunto. Apesar de poder ser entendido como sinonímia de “contexto situacional”, possui sentido mais amplo que envolve o ambiente em que está inserido além do conhecimento do usuário sobre o uso da língua. Contexto que se torna mais complexo quando se está em movimento, como é o caso do novo usuário móvel (FORESTI; VARVAKIS; GODOY VIERA, 2016, p. 4).

Nota-se a questão do usuário móvel, apontando para um condicionante de usuário múltiplo e variado dentro de um SI, em diversos contextos de uso específicos e entrelaçáveis. Cabe lembrar que a composição contextual de uso da informação é complexa, havendo intersecções e aproximações entre essas classificações (JOHNSON, 2003).

Em visão complementar, Choo (2003) nos apresenta três classificações para o contexto de busca, envolvendo questões específicas do usuário. São elas, a cognição, a emoção e a situação. Para o autor,

[...] a informação e o *insight* nascem no coração e na mente dos indivíduos, e que a busca e o uso da informação são um processo dinâmico e socialmente desordenado que se desdobra em camadas de contingências cognitivas, emocionais e situacionais (CHOO, 2003, p. 66).

A cognição tem a ver com a demanda lógico-operacional de um determinado usuário. O usuário demanda resolver um problema e se apropria da informação que entende que pode ajudá-lo para tal (CHOO, 2003). Belkin (1980) indicou que as pessoas buscam preencher lacunas de conhecimento para a resolução de problemas, chamando essa condição de estado anômalo de conhecimento. Fatores ambientais e educacionais têm peso no estado anômalo de conhecimento, normalmente, ligados

ao desenvolvimento comunicacional e educacional do usuário. Choo (2003) reforça este ponto dizendo que, pessoas

[...] em situações problemáticas, que buscavam informações, sentiam a inadequação de seu estado de conhecimento: "as inadequações de um estado de conhecimento podem ser de vários tipos, como lapsos ou falhas, incertezas ou incoerências, que só têm em comum uma sensação de erro". Belkin denominou-o estado anômalo de conhecimento. A hipótese do estado anômalo de conhecimento implica que os que buscam a informação são quase sempre incapazes de especificar suas necessidades, já que não conseguem expressar prontamente o que não sabem ou o que está faltando (CHOO, 2003, p. 80).

Este posicionamento se assemelha com o apresentado por Talja, Heidi e Tarja (1999) sobre os contextos extralinguísticos. Nota-se que a composição do arcabouço de conhecimentos de um usuário é um fator contextual que pode afetar a forma como a RI funciona. De fato, um dos fatores para a determinação da relevância da informação para um usuário está relacionado com seu conhecimento prévio sobre a mesma (HJØRLAND, 2010). Assim, a captação e compreensão do contexto cognitivo pode auxiliar no aprimoramento de sistemas de informação digitais.

A questão emocional tem a ver com o estado emocional/sentimental de um usuário no momento da RI, sendo, também, relativa a condições educacionais ao longo da vida do indivíduo. Essa questão pode impactar diretamente ou indiretamente na busca da informação de um usuário. No entanto, sua condição subjetiva necessita de ferramentas complexas para sua captação e uso, demandando, assim, fatores aproximados, ou atributos de representação, para a sua utilização.

O último ponto de classificação da informação, do ponto de vista do usuário, abordado por Choo (2003), é a situação. A situação envolve questões ambientais de espaço – iluminação, temperatura, conforto (ADAMS; SCHILIT; WANT, 1994) –, questões relativas a posicionamentos socioculturais, bem como, a experiência prévia que um determinado usuário tem sobre um determinado conjunto de informações. Em parte, o contexto situacional está ligado tanto com o estado anômalo de conhecimento quanto com a capacidade de um usuário em determinar se um código ou documento pode ser apropriado ou não. A classificação extralinguística de Talja, Heidi e Tarja (1999), assim como, a classificação situacional proposta por Choo (2003), engloba fatores linguísticos, emocionais, sociais e de conhecimento.

Neste caminho, Johnson (2003) apresenta duas possíveis classificações adicionais, com foco na forma de captação. Essas classificações são: a abordagem contingencial e a abordagem de enquadramento. De acordo com o autor as abordagens contingenciais “vão além da enumeração de fatores em uma situação para especificar os ingredientes ativos em um contexto e sua relação com os processos” ( JOHNSON , 2003, p. 740),, correlacionando estes fatores com as informações recuperadas.

O autor defende que as condições contextuais devem ser cuidadosamente consideradas em relação aos atributos contextuais ao buscar e recuperar informações. Ele reconhece que existem relações exclusivas e peculiares entre estes fatores, e que a compreensão de suas interações é fundamental para garantir que a informação recuperada seja relevante e útil para o usuário. Em outras palavras, a consideração cuidadosa do contexto e dos atributos contextuais é essencial para garantir a eficácia da recuperação de informações.

A abordagem contingencial melhora a compreensão da situação contextual, estabelecendo conexões mais fortes entre os elementos em um cenário de RI. Essa abordagem permite uma associação mais precisa entre os critérios de busca e as informações recuperadas, levando em consideração as particularidades do contexto em que a busca está sendo realizada. Em outras palavras, a abordagem contingencial considera a natureza contingente da recuperação de informações, adaptando-se às condições e às características específicas de cada situação para obter resultados mais eficazes.

Fry e Smith (1987), em contraponto, indicaram que os sistemas contingenciais têm problemas de equifinalidade, o que significa que é necessário ter vários cenários para atender todas as possibilidades de correlação entre busca e recuperação. A fim de mitigar esta problemática, Goffman (2012) sugere que padrões sejam estabelecidos entre relações contextuais e RI. Esta abordagem é chamada de enquadramento, trata de *frames* ou quadros. Para Goffman (2012), quadros representam um recorte de vivência de um indivíduo específico, ou usuário, caracterizando-o e caracterizando suas ações no tempo e no espaço. Apesar de identificar condições específicas dos usuários, dando um panorama substancial do indivíduo, Johnson (2003) afirma que os quadros limitam o contexto em faixas muito particulares, não tendo capacidade de generalização contextual.

Há, dessa forma, um ponto que requer análise sobre como estes atributos devem ser coletados, a fim de dar um panorama razoável sobre os contextos de um usuário. Sabe-se que a generalização de condições em sistemas com a escala de ambientes de grandes volumes de dados pode ser deletéria, no sentido de classificar indivíduos diferentes em grupos similares (O'NEIL, 2021; ZUBOFF, 2021). Atuar nessa seara pode ser interessante, tendo em vista a demanda de particularização e caracterização de um usuário em sistemas de informação.

Além disso, é importante acompanhar, dada a mudança do perfil dos usuários ao longo do tempo (ARAÚJO, 2017), os relacionamentos possíveis entre contextos de uso e RIs. Isso é importante, pois os usuários, como atores humanos e sociais, mudam ao longo tempo, dependendo do momento social, cultural e intelectual em que estão envolvidos. Usuários, como indivíduos, atuam como instituições com conexões de interesses que podem ser desfeitas de forma muito rápida e, por vezes, sem razão aparente (BAUMAN, 2001).

A consideração de um *frame* atualizável de usuário pode ser interessante como meio de captação de contextos específicos de um usuário ao longo do tempo. Também, ao relacionar *frames* de um indivíduo com a contingência entre busca e informação, a RI pode ocorrer de forma mais particularizada. A consideração da aplicação dessas abordagens para o projeto de SIs depende de sistemas de *hardware* e *software* robustos, dada a necessidade de particularização dos processos de busca e RI, e sua constante atualização.

Wersig e Windel (1985), em complemento a questão do usuário e sua relação com o meio, indicam que as ações do usuário, sejam elas técnicas ou mentais, são voltadas para seu objetivo, relacionados a um problema específico, do ponto de vista informacional. Isso explicita a observação de atributos no momento de resolução de problemas, passíveis de serem entendidos como atributos contextuais de uso. São eles: a situação atual do organismo, os estados passados do organismo, o sistema de preferências e o sistema de potencialidades. Rabello (2017, p. 126) as detalhou do seguinte modo:

- a) Situação atual do organismo: situação atual do usuário (desejos e aspirações);
- b) Estados passados do organismo: história pessoal do indivíduo;
- c) Sistema de preferências: valores, opiniões, atitudes, dentre outros aspectos;
- d) Potencialidades: cognitivas, afetivas, estéticas, dentre outras (RABELLO, 2017, p. 126).

As abordagens apresentadas não esgotam o leque de pesquisas existentes sobre contexto, mas apresentam um panorama inferencial do caminho que pesquisas nesse tema se dão. Nota-se que existem certas aproximações nas abordagens apresentadas, porém, sem consenso ou teoria unificada. Isso dificulta o desenvolvimento de um modelo conceitual definitivo de SI, orientado ao contexto de uso; porém, sua concepção não é impossibilitada. No Quadro 3 estão sumarizadas as abordagens estudadas, apresentando os atributos considerados em cada uma delas

**Quadro 3 - Atributos contextuais levantados na elaboração do referencial teórico**

<b>Autores</b>	<b>Atributos</b>	<b>Características dos Atributos</b>
NBR 9241-11	usuário, tarefa, equipamento e ambiente	Considera, respectivamente, o usuário e seu atributos específicos (linguagem, historicidade etc.); as ações frente ao sistema; os padrões físicos do sistema e o ambiente físico e sociocultural em que o sistema atua
Talja, Heidi e Tarja (1999)	objetificada e interpretativa	Consideram, respectivamente, atributos taxonômicos independentes; e atributos observados na mediação entre usuário e sistema
Foresti, Varvakis e Godoy Viera (2016)	linguístico/verbal, situacional e extralinguístico	Consideram, respectivamente, atributos de ordem linguística e simbólica, com relação ao idioma e estruturas de comunicação, ao conhecimento prévio de seu usuário e seu ambiente físico e sociocultural
Johnson (2003)	contingencial e enquadramento	Considera, respectivamente, atributos taxonômicos que relacionam o usuário e informação demandada e a possibilidade de criação de quadros padronizados para cada padrão contextual observado

continua...

conclusão...

**Quadro 3 - Atributos contextuais levantados no estudo de referencial teórico**

Wersig e Windel (1985)	técnicas e mentais	Consideram a situação atual do organismo, os estados passados do organismo, seu sistema de preferências e potencialidades
Choo (2003)	cognitivo, emocional e situacional	Considera, respectivamente, atributos ligados ao nível cognitivo de um usuário sobre um tema - ligado à comunicação -, sobre seu estado sentimental e emocional e sobre sua condição física, local, temporal, sociocultural e de conhecimento prévio de informação sobre um tema

Fonte: Elaborado pelo autor com base em Talja, Heidi e Tarja (1999), Foresti, Varvakis e Godoy Viera (2016), Choo (2003), Adams, Schilit e Want (1994), Henrique, Nassif e Venâncio (2007), Johnson (2003), Fry e Smith (1987) e Goffman (2012).

De maneira geral, a revisão revelou que existem três campos de atenção sobre atributos contextuais relevantes para a RI: sociocultural, linguístico/simbólico/interpretativo e cognitivo. O campo sociocultural abrange os atributos relacionados à formação do sujeito, sua interação com o outro e com o meio ambiente, bem como, os aspectos históricos, culturais e regionais que influenciam sua relação com a informação. Estes atributos são importantes para perfilar usuários, já que suas demandas informacionais podem variar de acordo com sua formação sociocultural e sua capacidade de apropriação da informação.

O campo linguístico/simbólico/interpretativo diz respeito à forma como um indivíduo percebe o mundo ao seu redor e se comunica, envolvendo aspectos estéticos e subjetivos, que influenciam na seleção e apresentação da informação. Estes atributos são influenciados pela formação sociocultural do usuário, mas também, possuem especificidades relacionadas à sua vivência. Já o campo cognitivo está relacionado à linguagem e ao conhecimento prévio do usuário, sendo mais lógico do que estético, influenciado pela comunicação do usuário, como idioma, jargões e regionalismos. Todos estes campos são importantes para aprimorar o uso dos

sistemas de RI e a apresentação da informação de forma mais adequada aos usuários.

## **2.7 Considerações gerais sobre o estado da arte de sistemas de informação digitais**

A evolução dos sistemas de informação, impulsionada por tecnologias como aprendizado de máquina e *chatbots*, transformou os usuários em produtores e disseminadores ativos de informação e conhecimento. Esse novo paradigma traz implicações relevantes para a gestão de sistemas de informação e redefine o conceito de sujeito informacional.

Essa transformação, contudo, enfrenta desafios cruciais relacionados à falta de identificação prévia do usuário nos sistemas digitais. Este obstáculo compromete seriamente a eficácia da recuperação da informação. Por outro lado, a mudança oferece oportunidades para tornar os processos de recuperação de informação mais personalizados e eficientes. Assim, a importância de considerar a perspectiva e as necessidades do usuário torna-se ainda mais crítica para o desenvolvimento e avaliação de SIs.

Neste contexto, a TI busca resolver essa questão utilizando ferramentas de *Machine Learning*, como o GPT, que se baseia em algoritmos de redes neurais artificiais para fornecer respostas em linguagem natural, para as perguntas feitas em ambientes digitais. Embora essas ferramentas possam resolver parte do problema de RI, elas possuem uma lógica própria, perpetuando condições de alienação de grupos de usuários que não fazem parte de suas estruturas de treinamento.

As tecnologias de GPT apresentam algumas problemáticas que também são observadas em SRIs digitais clássicos. O GPT apresenta desafios, tais como vieses, falta de transparência, manipulação de informações e dependência de grandes conjuntos de dados. É fundamental ter em mente essas questões ao utilizar o GPT, ou outras tecnologias de *Machine Learning*, para buscar maneiras de minimizar seus efeitos negativos e garantir o uso ético e responsável dessas ferramentas.

Para avançar na solução das questões de recuperação de informação, é essencial não somente considerar a evolução tecnológica, mas também aprofundar outras questões relativas ao usuário. Embora as ferramentas de ML e *chatbots* possam resolver parte do problema, precisamos reconhecer que ainda há muito a ser

aprimorado neste tipo de sistema. Por isso, é fundamental adotar uma abordagem que integre o olhar sobre o usuário com as novas tecnologias propostas, visando à resolução efetiva e responsável dessas questões.

Neste sentido, é fundamental considerar o contexto de uso da informação e o perfil do usuário, pois essas informações podem trazer uma visão mais qualificada do usuário. Ao levar em conta o contexto de uso, pode-se entender as necessidades específicas de cada usuário e, assim, oferecer informações mais relevantes e personalizadas. Além disso, ao considerar o perfil do usuário, podemos garantir que as ferramentas de RI estejam atendendo as necessidades de todos os usuários, evitando, assim, a perpetuação de condições de alienação de grupos que não fazem parte das estruturas de treinamento dessas ferramentas.

Assim, é preciso considerar que, embora as ferramentas de aprendizado de máquina possam trazer melhorias significativas para a gestão de sistemas de informação, não são a solução completa. É importante ter uma abordagem holística, que inclua tanto as tecnologias avançadas como a consideração dos usuários e suas necessidades. Dessa forma, será possível alcançar soluções de RI mais eficientes e justas.

## **CAPÍTULO 3 – PROCEDIMIENTOS METODOLÓGICOS**

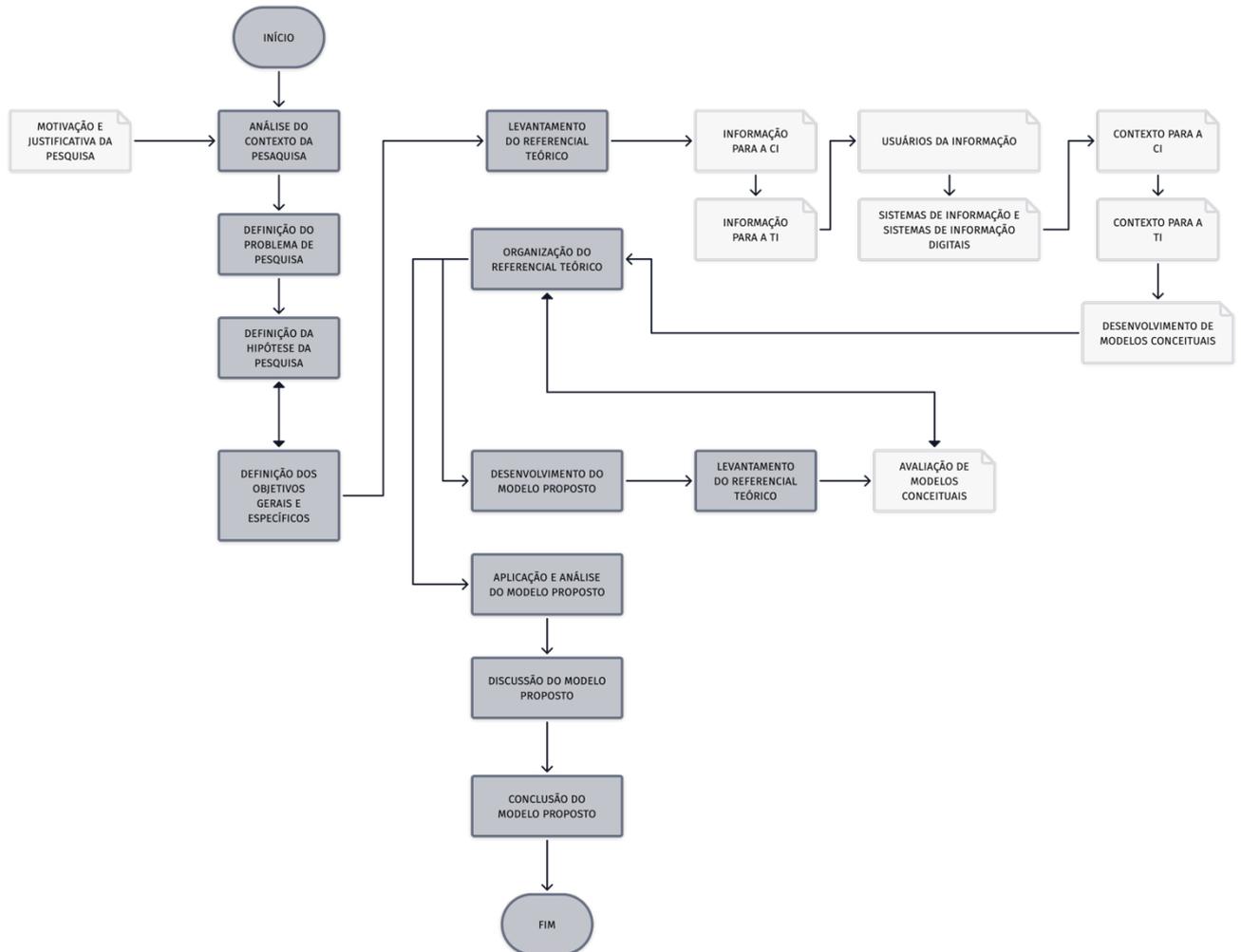
Este trabalho teve como objetivo principal propor um modelo que utilize elementos do contexto de uso para aprimorar as respostas fornecidas por um SRI utilizado em ambientes informacionais digitais. Os processos de recuperação de informações enfrentam diversos desafios e busca-se melhorar a eficiência e relevância das respostas geradas pelo SRI por meio da consideração do contexto de uso.

Esta pesquisa possui um caráter exploratório, no qual se busca levantar aspectos relacionados aos ambientes digitais informacionais e aspectos relacionados ao contexto de uso da informação. A abordagem adotada envolve a realização de um levantamento bibliográfico e estudo de caso, com o objetivo de mapear as condições de manifestação destes objetos de pesquisa. Neste sentido, a pesquisa explora a coleta de informações sobre um determinado objeto ou conjunto de objetos, delimitando assim um campo de trabalho e identificando as circunstâncias em que esses objetos se manifestam, conforme apontou Severino (2013) sobre pesquisas exploratórias.

A hipótese subjacente é que este contexto pode fornecer informações valiosas para melhorar as respostas fornecidas pelos sistemas de recuperação de informações. Estas informações estão relacionadas a três características gerais da informação: sua forma, conteúdo e acessibilidade, tanto física quanto cognitiva. O estudo é baseado em revisão da literatura existente, além da proposição de um modelo para a recuperação de informações com base no contexto do usuário. Além disso, é apresentada uma possibilidade de aplicação desse modelo.

Para o desenvolvimento da investigação, considerada a hipótese, foi adotado o seguinte fluxo, conforme fluxograma da Figura 12.

**Figura 12 - Fluxograma de desenvolvimento do trabalho**



Fonte: Elaborado pelo autor.

As etapas são descritas como seguem, em três momentos:

Estudo de Literatura e Preparação do Referencial Teórico: nesta etapa foram consultados repositórios de teses e dissertações, como os da USP, Unicamp, PUC, entre outros, bem como, bases de dados, como a Base de Dados Referencial de Artigos de Periódicos em CI (BRAPCI), *Web of Science* e Elsevier, a fim de dar base teórica e referencial para o desenvolvimento do modelo conceitual. Também, buscou-se situar o trabalho dentro do campo da CI, justificando sua pertinência e apresentando seu grau de originalidade.

Os principais autores utilizados no desenvolvimento do trabalho são apresentados no Quadro 4, organizados pelos seus temas gerais e descritores, utilizados para sua identificação nos repositórios supracitados:

Quadro 4 - Autores e temas utilizados no trabalho

Tópico	Descritores	Principais Referências
SISTEMAS DIGITAIS	Dados Massivos, <i>Business Intelligence</i> , <i>Data Value</i> , <i>Data Extraction</i>	BARLOW, M. <b>The Culture of Big Data</b> . Sebastopol: O'Reilly Media, 2013.
		BOYD, D.; CRAWFORD, K. Critical Questions for <i>Big Data</i> : Provocations for a cultural, technological, and scholarly phenomenon. <b>Information, Communication &amp; Society</b> , v. 15, n. 5, p. 662–679, maio 2012.
		CHEN, H.; CHIANG, R.; STOREY, V. Business Intelligence and Analytics: From <i>Big Data</i> to Big Impact. <b>MIS Quarterly</b> , v. 36, n. 4, p. 1165–1188, 2012.
		DAVENPORT, T. <b>Big Data at Work</b> : Dispelling the Myths, Uncovering the Opportunities. Boston: Harvard Business Review Press, 2014.
		DE MAURO, A.; GRECO, M.; GRIMALDI, M. A formal definition of Big Data based on its essential features. <b>Library Review</b> , v. 65, n. 3, p. 122–135, jan. 2016.
		EKBIA, H. et al. Big Data, Bigger Dilemmas: A Critical Review. <b>Journal of the Association for Information Science and Technology</b> , v. 66, n. 1, p. 1–34, 2015.
		MARQUESONE, R. <b>Big Data</b> : Técnicas e tecnologias para extração de valor dos dados. São Paulo: Casa do Código, 2016.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

USUÁRIOS DA INFORMAÇÃO	Usuário da Informação, Agente Informacional	GONZALEZ DE GÓMEZ, M. N.; RABELLO, R. Sujeito, agência e informação: tradição e leituras transversais. <i>In: Informação: agentes e intermediação</i> . Brasília: IBICT, 2017.
		LE COADIC, Y.-F. <b>A ciência da informação</b> . Brasília: Briquet de Lemos Livros, 2004.
		RABELLO, R. NOÇÕES DE SUJEITO EM MODELOS TEÓRICOS NA CIÊNCIA DA INFORMAÇÃO: do enfoque no sistema à consideração da agência em contexto. <b>informação e Sociedade</b> , v. 23, n. 3, p. 57–71, 2013a.
		RABELLO, R. Leituras sobre usuário e uso de informação na Ciência da Informação. <b>Perspectivas em Ciência da Informação</b> , v. 18, n. 4, p. 152–184, dez. 2013b.
		RABELLO, R. Sujeito e agência informacional: comportamento, prática e ação. <i>In: Informação: agentes e intermediação</i> . Brasília: IBICT, 2017.
		RABELLO, R.; GONZALEZ DE GÓMEZ, M. N. Agentes, intermediações e institucionalidades: apontamentos acerca de um mosaico interpretativo no campo informacional. <i>In: Informação: agentes e intermediação</i> . Brasília: IBICT, 2017.

continua...

continuação...

#### Quadro 4 - Autores e temas utilizados no trabalho

CONTEXTO	Contexto, <i>context</i> , <i>contexto- awareness</i>	ABOWD, G.; DEY, A. Towards a Better Understanding of Context and Context-Awareness. Lecture Notes in Computer Science. <b>Anais</b> [...]. Atlanta: Graphics, Visualization and Usability Center and College of Computing, jan. 1999.
		ADAMS, N.; SCHILIT, B.; WANT, R. Context-Aware Computing Applications. Workshop on Mobile Computing Systems and Applications. <i>In</i> : First workshop on mobile computing systems and applications. <b>Anais</b> [...]. Santa Cruz: IEEE Computer Society Press, dez. 1994.
		ABNT. <b>NBR 9241-11</b> : Requisitos Ergonômicos para Trabalho de Escritórios com Computadores Parte 11 – Orientações sobre Usabilidade. Rio de Janeiro: ABNT, 2011.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

CONTEXTO		COOL, C.; SPINK, A. Issues of context in information retrieval (IR): an introduction to the special issue. <b>Information Processing and Management</b> , v. 38, n. 1, p. 605–611, 2002.
		DEY, A. <b>Providing Architectural Support for Building Context-Aware Applications</b> . Tese. Georgia Institute of Technology, Atlanta, nov. 2000.
		DEY, A. Understanding and Using Context. <b>Personal and Ubiquitous Computing</b> , v. 5, n. 1, p. 4–7, 2001.
		NASSIF, M. E.; VENÂNCIO, L. S.; HENRIQUE, L. C. J. Sujeito, contexto e tarefa na busca de informação: uma análise sob a ótica da cognição situada. <b>DataGramZero - Revista de Ciência da Informação</b> , v. 8, n. 8, p. 1–12, out. 2007.
		TALJA, S.; HEIDI, K.; TARJA, P. The production of 'context' in information seeking research: a metatheoretical view. <b>Information Processing and Management</b> , v. 35, n. 1, p. 751–763, 1999.
CONCEITO DE INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO	Informação, Ciência da Informação, <i>Information Science</i> , <i>Information</i>	AWAD, E.; GHAZIRI, H. <b>Knowledge Management</b> . Patparganj: Dorling Kindersley, 2007.
		BARRETO, A. DE A. A questão da informação. <b>São Paulo em Perspectiva</b> , v. 8, n. 4, 1994.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

CONCEITO DE INFORMAÇÃO E CIÊNCIA DA INFORMAÇÃO	BELKIN, N. Anomalous states of knowledge as a basis for information retrieval. <b>Canadian Journal of Information Science</b> , v. 5, n. 1, p. 133–143, 1980.
	CAPURRO, R.; HJØRLAND, B. The concept of information. <b>Annual Review of Information Science and Technology</b> , v. 37, n. 1, p. 343–411, ago. 2003.
	DAVENPORT, T.; PRUSAK, L. <b>Conhecimento Empresarial</b> : como as organizações gerenciam o seu capital intelectual. Rio de Janeiro: Elsevier Science, 2003.
	SMIT, J. W. A informação na Ciência da Informação. <b>InCID: Revista de Ciência da Informação e Documentação</b> , v. 3, n. 2, p. 84–101, dez. 2012.
	SMIT, J. W.; BARRETO, A. DE A. Ciência da informação: base conceitual para a formação do profissional. <i>In: Formação do profissional da informação</i> . São Paulo: Polis, 2002. p. 9–23.
	TARAPANOFF, K. Informação, conhecimento e inteligência em corporações: relações e complementaridade. <i>In: Inteligência, Informação e Conhecimento</i> . Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2006. p. 19–35.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

PROCESSAMENTO DE DADOS	Inteligência Artificial, Machine Learning, Técnicas de Processamento de Dados, Sistemas Inteligentes, Artificial Intelligence	ARTERO, A. O. <b>Inteligência Artificial</b> . São Paulo: Livraria da Física, 2012.
		ASSIS, E.; GOUVÊA JR., M. Aprendizagem por Reforço com Rede Neural no Desenvolvimento de Jogos Digitais. <b>Instituto de Ciências Exatas e Informática</b> , v. 1, n. 1, p. 1–5, maio 2017.
		BITTENCOURT, G. <b>Inteligência artificial: Ferramentas e teorias</b> . Florianópolis, Santa Catarina, BRA: Editora da UFSC, 2006.
		BRAGA, A.; CARVALHO, A.; LUDERMIR, T. <b>Redes Neurais Artificiais: Teorias e Aplicações</b> . Rio de Janeiro, Rio de Janeiro, BRA: LTC, 2000.
		NORVIG, P.; RUSSELL, S. <b>Inteligência Artificial</b> . Rio de Janeiro: Campus, 2004.
		VASWANI, A. <i>et al.</i> Attention Is All You Need. <b>31st Conference on Neural Information Processing Systems</b> , p. 1–15, 2017.
RECUPERAÇÃO DA INFORMAÇÃO	Recuperação da Informação, Information Retrieval, Information Retrieval Systems, Sistemas de Recuperação da Informação	FERNEDA, E. <b>Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação</b> . São Paulo: USP - Universidade de São Paulo, 2003.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

RECUPERAÇÃO DA INFORMAÇÃO	FERNEDA, E. <b>Introdução aos Modelos Computacionais de Recuperação de Informação</b> . Rio de Janeiro: Ciência Moderna, 2012.
	INGWERSEN, P. Cognitive perspectives of information retrieval interaction. <b>Journal of Documentation</b> , v. 52, n. 1, p. 3–50, 1996.
	INGWERSEN, P. Cognitive information retrieval. <b>Annual Review of Information Science and Technology</b> , v. 34, p. 3–52, 1999.
	INGWERSEN, P.; JÄVERLIN, K. <b>The Turn: Integration of Information Seeking and Retrieval in Context</b> . Berlim: Springer, 2005.
	KUHLTHAU, C. <b>Information search process</b> . Boca Raton: Encyclopedia of Library and Information Sciences, 1991.
	MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. <b>Introduction to Information Retrieval</b> . Cambridge: Cambridge University Press, 2008.
	VICKERY, A.; VICKERY, B. <b>Information Science in Theory and Practice</b> . Munique: K.G. Saur, 2004.
	WHITE, T. <b>Hadoop, the definitive guide: Storage and Analysis at Internet Scale</b> . Sebastopol: O'Reilly Media, 2015.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

RELEVÂNCIA DA INFORMAÇÃO	Relevância da Informação, Pertinência da Informação, Information Relevance, Informação Útil	BORLUND, P. The Concept of Relevance in IR. <b>Journal of the Association for Information Science and Technology</b> , v. 54, n. 10, p. 913–925, 2003.
		CAMPBELL, I.; VAN RIJSBERGEN, K. The ostensive model of developing information needs. <i>In</i> : P. <b>Proceedings of CoLIS 2, second international conference on conceptions of library and information science: Integration in perspective</b> . Copenhagen: Royal School of Librarianship, 1996. p. 251–268.
		COEIRA, E. W.; VICKLAND, V. Is relevance relevant? User relevance ratings may not predict the impact of Internet search on decision out- comes. <b>Journal of the American Medical Informatics Association</b> , v. 15, n. 4, p. 542–545, 2008.
		FOSKETT, D. J. A note on the concept of “relevance”. <b>Information Storage Retrieval</b> , v. 8, n. 1, p. 77–78, 1972.
		HARTER, S. P. Psychological relevance, and information science. <b>Journal of the American Society for Information Science</b> , v. 43, p. 602–615, 1992.
		HJØRLAND, B. The Foundation of the Concept of Relevance. <b>Journal of the American Society for Information Science</b> , v. 61, n. 2, p. 217–237, 2010.

continua...

continuação...

**Quadro 4 - Autores e temas utilizados no trabalho**

RELEVÂNCIA DA INFORMAÇÃO	MIZZARO, S. How many relevances in information retrieval? <b>Interacting with Computers</b> , v. 10, p. 303–320, 1998.
	REID, J. A New, task-oriented paradigm for information retrieval: Implications for evaluation of information retrieval systems. <i>In: Proceedings of CoLIS 3, third international conference on the conceptions of library and information science</i> . Lovke: Naklada Benja, 1999. p. 97–108.
	SARACEVIC, T. RELEVANCE: A Review of and a Framework for the Thinking on the Notion in Information Science. <b>Journal of the American Society for Information Science</b> , p. 321–343, 1975.

continua...

conclusão...

#### Quadro 4 - Autores e temas utilizados no trabalho

SISTEMAS DE INFORMAÇÃO	Sistemas de Informação, Sistemas de Recuperação da Informação, <i>Information Systems and IS</i>	CHOO, C. W. Como ficamos sabendo – um modelo de uso da informação. <i>In: A organização do Conhecimento</i> : como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Editora SENAC, 2003. p. 61–120.
		FUJINO, A. Acessibilidade informacional de PCD no contexto da lei de acesso à informação: desafios para estudo de usuários. <b>Informação em Pauta</b> , v. 2, p. 237-257, 2017.
		ARAÚJO, C. A. Á. Práticas informacionais: novo conceito para o estudo dos usuários da informação. <i>In: Informação</i> : agentes e intermediação. Brasília, Distrito Federal, BRA: IBICT, 2017.
		FIGUEIREDO, N. M. <b>Estudos de Uso e Usuários da Informação</b> . Brasília: IBICT, 1994.
		GUINCHAT, C.; MENO, M. <b>Introdução geral às ciências e técnicas da informação e documentação</b> . Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1994.
		SANZ CASADO, E. <b>Manual de estudios de usuarios</b> . Madri: Fundación Germán Sánchez Ruipérez, 1994.

Fonte: Elaborado pelo autor.

Desenvolvimento de um modelo conceitual para recuperação da informação em ambientes informacionais digitais: o presente trabalho adotou a abordagem de modelagem de sistemas para análise da utilização de contexto de uso da informação e contexto das características do usuário. A modelagem de sistemas de informação considera três etapas básicas: conceitual, física e lógica (BALCI, 2007). Cada uma dessas etapas envolve diferentes tipos de modelos e artefatos, tendo como objetivo final, a implementação de um SI funcional e eficaz que atenda aos requisitos do usuário.

Na etapa de modelagem, foi criado um modelo abstrato das ideias e conceitos do SI, sem considerar os aspectos técnicos. Na modelagem lógica, o modelo conceitual é transformado em uma estrutura mais organizada e coerente, que pode ser processada por um computador. Já na modelagem física, o sistema é implementado em um ambiente tecnológico específico, dando vida ao modelo lógico e o tornando funcional, pronto para ser utilizado como um sistema finalizado (BALCI, 2007).

Para o projeto em questão, buscou-se aplicar os conceitos de contexto de usuário e uso em sistemas de recuperação de informações digitais, com o objetivo de analisar seu potencial de aprimoramento, sem a intenção de executá-lo ou aplicá-lo em uma tecnologia específica, faz-se importante utilizar técnicas da primeira etapa de modelagem conceitual.

Utilizando técnicas de modelagem conceitual, foi possível aplicar e testar a implementação operacional do contexto de usuário e uso em processos de sistemas de recuperação de informação digitais. Essa abordagem permitiu avaliar em que medida a aplicação do contexto de usuário e uso melhora a conexão entre a busca e a recuperação da informação.

De fato, três foram os fatores considerados para a seleção da fase de modelagem conceitual como método para o desenvolvimento do presente trabalho:

- a) Os processos de trato da informação, como produção, armazenamento, busca, recuperação e organização, observando os ambientes em que essas atividades ocorrem. No caso de ambientes informacionais digitais, esta observação se mostrou difícil devido à escala e dinamicidade do ambiente, o que torna difícil a análise em tempo real. Para superar esta dificuldade, o desenvolvimento de

um modelo conceitual pode ajudar a criar um espaço de controle, em que a frequência de atualização pode ser determinada pelo pesquisador.

- b) Ao se concentrar, principalmente, em teorias e conceitos, a modelagem conceitual oferece uma abordagem livre para o desenvolvimento do sistema proposto, sem se preocupar com as amarras técnicas da implementação (BARUZZO *et al.*, 2009). Ao usar a linguagem dos modelos conceituais, este trabalho fornece, aos futuros projetistas de sistemas, resultados que podem ser utilizados, a partir de uma linguagem comum e de fácil compreensão.
- c) A modelagem conceitual possui técnicas estabelecidas de avaliação que não exigem o desenvolvimento lógico ou físico para avaliar a qualidade do modelo (BALCI, 2007; RECKER *et al.*, 2021). Dessa forma, ao usar a modelagem conceitual, de acordo com as práticas estabelecidas para sua construção, é possível validar e verificar o potencial de funcionamento de um sistema sem a necessidade de desenvolver sua implementação.

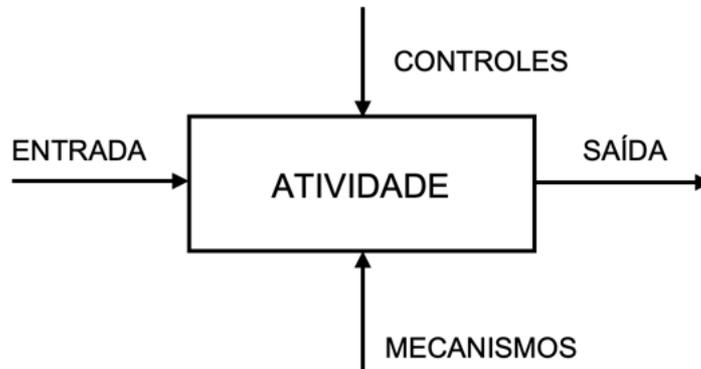
Para o desenvolvimento do modelo, foram utilizadas técnicas como o *Icam* *DEFinition for Function Modeling 0* (IDEF-0) e *use case*.

O IDEF-0 é especificamente utilizado para descrever as funções ou atividades que ocorrem em um sistema ou processo. Nesse modelo, as funções são representadas em diagramas que mostram como as entradas são transformadas em saídas por meio de processos intermediários. Esses diagramas ajudam a visualizar e entender as interações entre as funções, os recursos envolvidos e as restrições que afetam o sistema ou processo em questão.

Para compreensão do funcionamento do IDEF-0, entende-se que o elemento básico do diagrama consiste em um retângulo, que representa cada atividade ou processo. Este retângulo é acompanhado por quatro setas posicionadas ao seu redor. A primeira seta representa as entradas, ou seja, o que será transformado durante a atividade ou processo. A segunda seta representa os controles, que influenciam ou direcionam a maneira como o processo é executado, como *softwares* e regras. A terceira seta representa os mecanismos necessários para a realização da atividade, tais como pessoas, ferramentas, máquinas e equipamentos. Por fim, a quarta seta representa as saídas, que são os resultados gerados pela atividade e que podem ser

encaminhados para outros processos ou utilizados de maneira final por algum indivíduo. A Figura 13 apresenta este bloco de forma simplificada.

**Figura 13 - Bloco da estrutura básica do diagrama IDEF-0**



Fonte: Elaborado pelo autor, com base em Dašić *et al.* (2009).

Sobre o *use case*, ou estudo de case de uso, este é uma ferramenta de modelagem que descreve as interações entre um sistema e seus atores, que podem ser usuários humanos, outros sistemas ou entidades externas. Cada caso de uso representa uma funcionalidade específica do sistema e inclui um ator que desencadeia o processo e um conjunto de ações que ocorrem, detalhando como o sistema responde às ações do ator. Essa técnica é amplamente utilizada para documentar e compreender os requisitos funcionais de um sistema, fornecendo uma visão clara das interações entre os usuários e o software, o que é fundamental para o desenvolvimento e a validação de sistemas de software.

Exposição de uma possível aplicação do modelo proposto: a metodologia adotada nesta pesquisa consistiu em mostrar as possibilidades de aplicação e uso do modelo proposto frente à uma operação de “mundo real”. Para isso, utilizou-se como referência o SRI digital da BRAPCI, que é uma base de dados referencial de artigos de periódicos em CI.

A BRAPCI é uma base de dados referencial online que reúne informações bibliográficas e resumos de artigos científicos relacionados à área de CI. Ela desempenha um papel importante como uma ferramenta para pesquisadores e

profissionais interessados nessa área, oferecendo acesso a um vasto acervo de conhecimento científico.

Trata-se de uma base de dados referencial que estabelece conexões entre documentos e artigos presentes em várias outras bases de dados. Esta atua como uma interface centralizada que disponibiliza os endereços dos documentos, permitindo o acesso facilitado a estes documentos. É importante ressaltar que as bases de dados referenciais contêm apenas as referências dos documentos, podendo incluir artigos de revistas, livros, dissertações, teses e outros tipos de conteúdo, com ou sem resumos.

O objetivo principal foi demonstrar as possibilidades de utilização do modelo proposto. Para isso, realizou-se uma aplicação prática do modelo na operação real da BRAPCI, servindo como referência. Essa aplicação permitiu identificar as vantagens e desvantagens potenciais do modelo em termos de uso e aplicação no contexto da CI. Além disso, possibilitou uma análise cuidadosa das melhores formas de integração do modelo com uma base de dados real, considerando os desafios que podem surgir neste processo.

Para avaliar o modelo apresentado neste estudo, foi adotada a análise SWOT como método de análise. Esta abordagem proporcionou uma avaliação abrangente, considerando as Forças, Fraquezas, Oportunidades e Ameaças do modelo. Tal estratégia permitiu uma análise minuciosa e completa do modelo, destacando seus aspectos positivos, áreas de melhoria, oportunidades de aplicação e possíveis desafios. Isso proporcionou uma visão balanceada do modelo proposto no estudo.

## **CAPÍTULO 4 – MODELO PARA RECUPERAÇÃO DE INFORMAÇÕES**

Neste capítulo, intitulado "Modelo para Recuperação de Informações", é apresentado o modelo desenvolvido neste trabalho para a recuperação de informações em ambientes informacionais digitais. Conforme mencionado, o modelo utiliza atributos do contexto de uso da informação como termos vinculadores entre as expressões de consulta e busca da informação e documentos armazenados em ambientes digitais.

Inicialmente, são apresentadas considerações gerais para a proposição do modelo. Em seguida, o modelo é apresentado de maneira abrangente, explicando seu funcionamento geral em relação ao usuário, ao contexto de uso da informação e ao SI.

Posteriormente, o modelo é detalhado em termos de funcionamento e operação, fornecendo uma visão mais aprofundada de como ele é implementado e poderá ser aplicado. Este detalhamento inclui aspectos técnicos e práticos, ressaltando a interação entre os diferentes elementos do modelo. A ênfase é dada na descrição detalhada do modelo, relacionando-o com o referencial teórico que embasa o trabalho.

Por fim, o modelo é submetido a avaliação de possibilidade de uso. O objetivo foi indicar como o modelo conceitual proposto poderá ser utilizado em situação de vida real. Através dessa comparação, são identificadas vantagens, desvantagens e possíveis aprimoramentos dos processos de recuperação de informações.

#### **4.1 Considerações gerais da proposição do modelo**

Conforme apontado por Lima e Campos (2022), o processo de recuperação da informação pode ser organizado em quatro etapas: entrada, processamento, saída e avaliação. Simplificadamente, estas etapas envolvem a seleção e obtenção das informações, organizando-as, classificando-as e indexando-as para que estejam relacionadas às necessidades de informação dos usuários, conforme expressas em suas buscas, de forma a disponibilizar informações relevantes para recuperação.

Após ação de busca e recuperação da informação, o usuário avalia a informação recuperada e o sistema, indicando se a informação foi relevante e

apresentando críticas ao sistema. Todas estas ações pressupõem conhecimento prévio do usuário que utilizará o sistema. As ações de busca e recuperação da informação compreendem os processos de procurar e obter informações relevantes para atender a uma necessidade informacional específica. Neste processo, o usuário realiza uma busca utilizando palavras-chave, frases ou outros critérios relevantes para localizar as informações desejadas. Em seguida, as informações são encontradas e apresentadas ao usuário, possivelmente de forma relevante e compreensível.

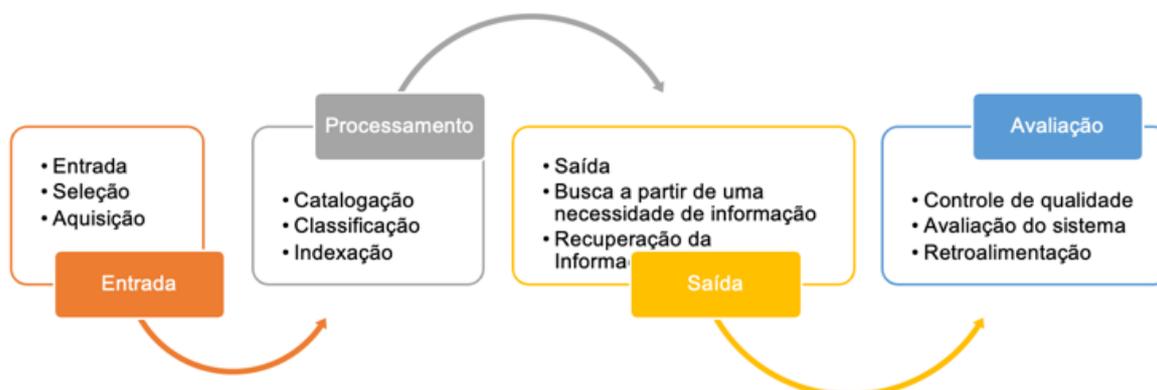
Para fornecer informações relevantes a um usuário, é essencial compreender seu perfil, incluindo suas características, preferências, interesses e comportamentos em relação ao uso da informação, bem como os contextos em que o usuário pode usar as informações (FROHMAN, 2008). Estes aspectos do perfil do usuário desempenham um papel importante na recuperação de informações que sejam relevantes e pertinentes o usuário em questão.

Ao realizar atividades de seleção e aquisição, os gestores dos sistemas de informação identificam, com base no perfil do potencial usuário do sistema e suas possíveis utilizações, quais documentos, arquivos e recursos podem conter informações relevantes, visando formar um acervo mais adequado às suas características. Isso difere-se dos SIs em ambientes digitais, que coletam toda a produção possível, sem seleção prévia. A representação dos documentos desempenha um papel fundamental na recuperação da informação, pois analisar individualmente cada documento em relação ao seu conteúdo e tema para responder a consultas específicas é praticamente impossível e demorado. Ao representar um documento ou arquivo usando termos e palavras-chave, é possível estabelecer vínculos entre os termos da consulta e os arquivos armazenados, facilitando a recuperação de informações relevantes (LIMA; CAMPOS, 2022).

Após a recuperação da informação, é importante que o usuário avalie a sua relevância e pertinência. Esta avaliação serve como base para a melhoria contínua do processo de recuperação da informação, pois indica as melhores relações entre documentos, arquivos e indexadores.

A Figura 14, retirada de Lima e Campos (2022), explicita este processo de forma geral.

**Figura 14 – Fluxo informacional em um SRI, conforme Lima e Campos (2022)**



Fonte: Lima e Campos (2022, p. 3).

Estas ações estão alinhadas com o explicitado por Smit e Barreto (2002) ao explicarem os fluxos informacionais de estoques informacionais, indicando que a informação, após sua criação, deve ser selecionada e armazenada a partir de uma lógica, organizada no sentido de ser recuperada por um usuário específico que, a depender da relevância e pertinência da informação recuperada, poderá, eventualmente, assimilá-la e se apropriar dela.

Ingwersen (1999) propôs um modelo de recuperação da informação que enfatiza a importância da organização da informação para atender a problemas e objetivos específicos dos usuários. De acordo com este modelo, é essencial realizar essa organização de forma antecipada, o que requer estruturas de representação e indexação da informação para facilitar sua recuperação.

Além disso, Ingwersen (1999) ressaltou que este modelo deve se adaptar constantemente para atender às necessidades em evolução dos usuários, que mudam ao longo do tempo e variam de acordo com seu nível atual de conhecimento. Esta adaptação contínua é baseada na avaliação do usuário. O proposto por Ingwersen (1999) está em consonância com o conceito de *feedback* de relevância.

O *feedback* de relevância é uma técnica utilizada em sistemas de recuperação de informação para melhorar a precisão das consultas ao longo do tempo. Este método ajusta automaticamente as consultas com base na interação do usuário, utilizando informações sobre quais documentos foram considerados relevantes ou não relevantes. Diversas abordagens têm sido aplicadas para integrar a avaliação de

relevância em sistemas de recuperação de informação, variando desde métodos tradicionais, como o algoritmo de Rocchio que reformula a consulta com base em uma combinação linear de vetores, até abordagens mais modernas que empregam modelos de linguagem pré-treinados e redes neurais para a reclassificação dos documentos recuperados (BAUMGÄRTNER et al., 2022).

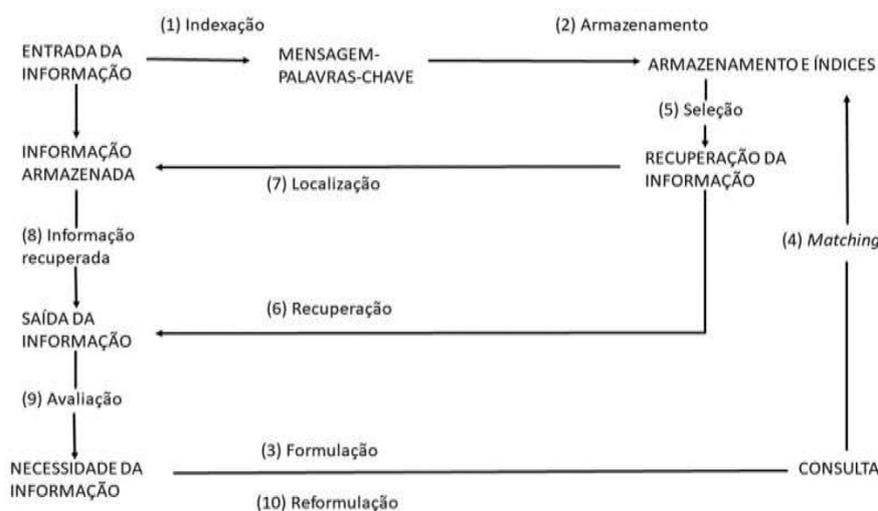
O uso de *feedback* de relevância potencializa o desempenho dos sistemas de recuperação de informação ao permitir ajustes na consulta e nos critérios de classificação. O feedback pode ser incorporado tanto como um parâmetro para atualizar o modelo de consulta, quanto para reordenar documentos em uma fase subsequente de recuperação, como observado em métodos de reclassificação por rede neural (BAUMGÄRTNER et al., 2022). Essas técnicas oferecem avanços significativos sobre métodos puramente lexicais, e a aplicação de aprendizado com poucos exemplos (*few-shot learning*) permite uma maior flexibilidade no uso de feedback explícito para refinar consultas e melhorar a precisão na recuperação de documentos relevantes (BAUMGÄRTNER et al., 2022).

Neste mesmo sentido, Vickery e Vickery (2004) indicaram também que os processos esperados de um SRI incluem, para recuperação da informação, seleção prévia do que se deseja armazenar, organizando, classificando e representando essa informação, por meio de índice, a fim de garantir a localização da informação no estoque a partir da consulta, para que essa seja recuperada com fins de atender uma necessidade informacional específica de um usuário. Após a recuperação de informações, há processo de avaliação por parte do usuário, assim como observado em Ingwersen (1999) e Lima e Campos (2022) para aprimorar os processos futuros de consulta e recuperação da informação.

O mesmo pode ser visto em Ferneda (2012) que indicou em seu modelo de RI a questão da representação da informação, necessária para alinhamento da função de busca e recuperação da informação. Em complemento, Choo (2003) indicou a necessidade de avaliação da informação recuperada por parte do usuário, com intuito de dar retorno para melhoria dos processos de RI. Estes exemplos reforçam o apontado por Lima e Campos (2022) sobre a necessidade prévia de seleção e organização do estoque informacional, aplicando políticas de representação da informação e indexação dos meios informacionais para, a partir das consultas, vincular

busca e recuperação da informação. Isso, visando recuperar informações úteis para um perfil de usuário específico. Estes processos podem, de forma geral, ser observados e entendidos como atributos de sistemas clássicos de informação. A Figura 15, retirada de Vickery e Vickery (2004), sumariza este processo.

**Figura 15 - Fluxo informacional em um SRI, conforme Vickery e Vickery (2004)**



Fonte: Vickery e Vickery (2004, p. 118).

Em SIs digitais, no entanto, nos quais a velocidade e o volume de produção de dados são altos e não há uma seleção prévia ou análise dos dados coletados e armazenados, a atividade de representação e indexação se torna inviável (EK BIA *et al.*, 2015). Mesmo que a produção de dados fosse controlada, estes sistemas coletam informações sobre diversas temáticas, sem serem direcionados a um perfil de usuário específico. Neste contexto, a hipótese deste trabalho considera que o contexto de uso da informação pode ajudar a solucionar este problema, relacionando tanto o usuário quanto o seu problema informacional aos arquivos e documentos potencialmente relevantes armazenados em um grande volume de dados.

Tal produção ocorre em virtude das mudanças observadas nos regimes informacionais, dadas as novas TDICs (GONZALEZ DE GÓMEZ, 2012). Estes novos regimes, propiciados por exemplo por tecnologias mobile, faz com que usuário passe também a produzir documentos e arquivos com informações passíveis de recuperação em grande quantidade (RABELLO, 2013b), de forma desestruturada e

com pouca caracterização. Isso cria dificultadores no sentido de (1) selecionar o que deve ser armazenado, (2) a forma de classificação, (3) a forma de representação e indexação dos documentos e arquivos a partir das informações estocadas.

Para tentar mitigar estas barreiras, são utilizados métodos e tecnologias que tentam organizar essa produção para fins de uso futuro. Como visto, estes métodos e técnicas visam organizar este grande volume de dados em bancos de dados do tipo NoSQL para consultas futuras, a partir de técnicas de aprendizado de máquina alicerçadas em algoritmos computacionais estatísticos, que tentam vincular termos de busca e metadados de arquivos e documentos salvos (DE MAURO; GRECO; GRIMALDI, 2016; WHITE, 2015).

Esses processos ainda são insatisfatórios em parte significativa de consultas feitas a estes sistemas, tendo em vista que demandariam de controle prévio do que armazenar para pleno funcionamento de vínculo entre termos de busca e metadados. A revocação neste tipo de sistema é sempre bastante alta, tendo em vista que a especificidade dos metadados retirados dos documentos, utilizando as técnicas já existentes, é baixa.

Para garantir que todos os aspectos relevantes sejam considerados na classificação de documentos em tais espaços, as regras de geração de metadados, que utilizam técnicas como recuperação de conhecimento em textos, frequentemente consideram a exaustividade máxima possível. Isso pode levar à inclusão de termos que têm pouca ou nenhuma relevância para determinados contextos de uso da informação contida no documento, resultando em consultas futuras que não correspondem às necessidades informacionais dos usuários.

Apesar do avanço de técnicas avançadas, como o GPT, apresentadas inicialmente por Vaswani *et al.* (2017), que utiliza redes neurais artificiais de atenção para melhorar o processamento de dados em sistemas, essas técnicas ainda podem encontrar limitações ao lidar com grandes volumes de dados e informações divergentes provenientes de diferentes contextos, especialmente quando aplicadas para aprimorar SRIs.

Nota-se um problema explícito de vinculação entre consulta e busca, tendo em vista a estrutura de organização e indexação dos documentos e arquivos armazenados neste tipo de ambiente. Uma representação deficitária da informação,

conforme apontado por Novellino (1996) e reforçado por Dahlberg (2006), também referenciado por Lima e Campos (2022), pode levar à uma recuperação da informação deficitária.

Ao considerar o problema de pesquisa proposto nesta tese, busca-se entender como é possível promover a recuperação de informações em ambientes informacionais digitais, tendo em vista que estes ambientes são diferentes dos “tradicionais” sistemas de informação. Dessa forma, apresenta-se uma forma de aprimoramento baseada na vinculação de termos de busca e consulta com os documentos e arquivos armazenados. Isso é feito considerando-se a hipótese apresentada, da utilização de atributos contextuais de uso da informação, com fins de melhor qualificar a informação recuperada para um usuário.

O desenvolvimento de sistemas tem uma história de uso de atributos contextuais para melhorar os processos. Em relação à usabilidade e à criação de produtos eficazes e eficientes, a norma NBR 9241-11 (ABNT, 2021) já reconhece a importância do contexto de uso de um sistema, produto ou serviço em seu design, como visto na figura 12. Isso significa que considerar o contexto em que o sistema será utilizado é fundamental para garantir sua usabilidade e efetividade. Barbosa e Silva (2010) e Lima e Campos (2022) também enfatizaram a questão de contexto de uso, tanto em projeto de sistemas computacionais quanto em projetos de SI, respectivamente.

Sabe-se, também, que as demandas informacionais mudam constantemente, bem como perfis de usuário, dada sua realidade e necessidades informacionais (FIGUEIREDO, 1994). Neste sentido, considerar o contexto de uso da informação para aprimorar estes SRIs em ambientes digitais, que abarca questões relativas ao usuário e sua demanda informacional, podem dar luz ao problema de pesquisa.

#### **4.2 Modelo proposto para recuperação da informação em ambientes informacionais digitais**

O objetivo principal deste trabalho consiste em propor um modelo que empregue elementos do contexto de uso, com o intuito de melhorar as respostas fornecidas por um SRI utilizado em ambientes informacionais digitais.

Neste sentido, é proposto um modelo baseado no contexto de uso da informação, que tem como objetivo caracterizar o usuário e suas necessidades informacionais em ambientes informacionais digitais a partir do exposto sobre sensibilidade ao contexto, visto em Dey (2001). Esta caracterização busca estabelecer possíveis conexões entre o estoque de informações e as demandas dos usuários, reduzindo o impacto da falta de identificação destes atores neste tipo de sistema.

Entende-se que, dada a característica de produção em grande volume e em grande velocidade de ambientes informacionais digitais, o processamento prévio de seus documentos e arquivos se torna tarefa desafiadora (EK BIA *et al.*, 2015). Desta forma, processar estes documentos e arquivos com objetivo de aplicar políticas específicas de indexação para a RI se torna tarefa virtualmente impossível. Sabe-se que a indexação e classificação documental prévia é necessária para garantir uma recuperação de informação mais alinhada com a necessidade de informacional de um usuário (LIMA; CAMPOS, 2022; NOVELLINO, 1996). A indexação é feita com base no potencial de relevância que um documento pode ter para um perfil específico de usuário. Ao não ser capaz de realizar este processo de forma prévia, estes ambientes informacionais carecem da capacidade de identificar as necessidades de seus usuários, tendo uma RI deficiente.

No contexto do modelo proposto, busca-se estabelecer uma conexão entre os atributos contextuais de uso da informação e os documentos ou arquivos armazenados em um repositório digital. Dado que o processo de representação dos documentos é desafiador neste tipo de ambiente, devido à falta de informações prévias sobre o usuário, o modelo propõe a vinculação posterior entre os documentos e os atributos contextuais por meio de "termos vinculadores". Estes termos vinculadores são gerados a partir da captura dos atributos contextuais socioculturais e cognitivos durante o processo de consulta e busca de informações, e funcionam como elementos físicos que possibilitam a conexão entre os documentos e os atributos contextuais.

O conceito de "termo vinculador" é oriundo da TI, se referindo a mecanismos ou componentes utilizados para estabelecer conexões entre diferentes elementos ou entidades de um sistema, sejam eles de operação ou documentais (TOCCI; WIDMER; MOSS, 2016). Estas conexões podem ser estabelecidas para facilitar a comunicação,

compartilhamento de informações, integração de dados ou coordenação de processos. Eles desempenham um papel fundamental na interoperabilidade e na interconexão de componentes de um sistema, permitindo que diferentes partes de um sistema se comuniquem e compartilhem dados de forma eficiente e eficaz (TOCCI; WIDMER; MOSS, 2016). Neste sentido, a utilização de “termos vinculadores” pode ser considerada para a conexão entre atributos contextuais e documentos, considerando que essas conexões são necessárias para a operação do sistema, e se assemelham a este dispositivo técnico já utilizado para o projeto de sistema pela TI. A sua utilização será detalhada adiante.

De forma simplificada, o BD do SI digital do modelo armazena todos os documentos e arquivos disponíveis que lhe são enviados, como atualmente são trabalhados os ambientes digitais. Durante o processo de armazenamento, os documentos e arquivos são submetidos a técnicas de aprendizado de máquina para identificação de palavras-chave e termos com maior frequência em seu conteúdo. A ideia não é descrever os documentos em detalhes como em um processo de indexação, mas sim, encontrar termos suficientes para uma recuperação inicial. Por isso, a descrição do documento para a recuperação inicial será feita com o maior grau de exaustividade que a ferramenta de aprendizado de máquina possa oferecer.

Ao acessar o SRI proposto no modelo, o usuário é solicitado a fornecer informações sobre seu perfil, incluindo detalhes relacionados à ocupação, localização de residência, escolaridade e outros atributos relevantes para sua classificação. Estas informações compõem o conjunto de atributos contextuais socioculturais do usuário. Como não é possível verificar a veracidade das informações apresentadas, três informações são coletadas automaticamente: localização *Global Positioning System* (GPS), data e horário da consulta, e tipo de dispositivo utilizado. As demais informações são obtidas por meio de um formulário preenchido pelo usuário.

Importante indicar que identificar informações falsas em formulários pode ser difícil por várias razões. Indivíduos que disseminam informações falsas podem demonstrar habilidades refinadas em manipulação e utilizar técnicas de engenharia social para se apresentarem como fontes legítimas de informação. Além disso, muitos formulários não possuem um processo sólido de verificação de identidade, o que dificulta a distinção entre informações verdadeiras e falsas. Ademais, diferentes tipos

de informações falsas podem ser fornecidos, como nomes, endereços e históricos profissionais falsos, o que requer métodos diferentes de detecção.

A verificação completa de cada informação em um formulário pode ser demorada e requer recursos consideráveis. Além disso, as técnicas de falsificação estão em constante evolução, tornando ainda mais desafiador identificar informações falsas. Neste sentido, entendendo que a presente pesquisa objetiva o aprimoramento dos processos de recuperação de informação em ambientes digitais, o modelo proposto não apresenta uma solução inicial para este problema, cabendo esta investigação às pesquisas futuras.

Sobre captação dos atributos contextuais de forma automática, este conceito reflete a ideia originalmente proposta por Abowd e Dey (1999) em relação aos sistemas sensíveis ao contexto. De acordo com Abowd e Dey (1999), o contexto de uso da informação, e conseqüentemente seu usuário, é inicialmente definido por atributos espaciais e temporais. Estes atributos, segundo os autores, possuem particularidades que ajudam a identificar o usuário da informação, considerando que diferenças no tempo e no espaço podem influenciar as características dos usuários. Portanto, esses atributos foram selecionados para a captura automática.

Outras informações relevantes sobre o usuário são coletadas por meio de um formulário. O sistema possui um mecanismo interno que descarta atributos contextuais que deixaram de ser relevantes para fortalecer as conexões entre os mapas contextuais e os documentos/arquivos. Espera-se que a coleta de informações atípicas por meio do preenchimento do formulário seja rara, e esses dados sejam posteriormente eliminados durante o processo de limpeza do sistema, ao cruzar os mapas contextuais mais frequentemente utilizados em consultas com perfis semelhantes que ocorrem com frequência.

De forma inicial, os campos do formulário devem englobar questões relacionadas à localização, experiência profissional, crenças religiosas e histórico acadêmico. Esses campos podem ser descritos da seguinte maneira:

- a) Ocupação;
- b) Religiosidade;
- c) Grau de instrução;
- d) Formação;

- e) Local de nascimento;
- f) Cidade de residência;
- g) Bairro.

A seleção destes campos é determinada conforme mencionado por Rabello (2017), que sugere que o contexto de um usuário pode estar relacionado com a situação atual do organismo (desejos e aspirações), estados passados do organismo (história pessoal do indivíduo), sistema de preferências (valores, opiniões, atitudes) e potencialidades (cognitivas, afetivas, estéticas), entre outros aspectos. No entanto, é necessário realizar estudos futuros, durante a implementação do modelo, para determinar quais campos devem ser removidos ou adicionados, de forma a impactar diretamente na construção dos vínculos entre consultas e documentos.

Subsequentemente a este processo de descrição, o usuário pode explicitar seu problema ou necessidade informacional, tanto com base em descritores quanto em linguagem natural. Após a explicitação do problema do usuário, o sistema identifica termos e repetições de palavras em sua busca, bem como expressões específicas utilizadas para realizar sua solicitação. Estes termos e conjuntos de palavras compõem o arcabouço de atributos contextuais cognitivos, tendo em vista que estão relacionados com questões relativas à idioma, descrição do problema e estado anômalo de conhecimento. Eventualmente, a depender da simplicidade dos argumentos de busca, estes valores podem ser desconsiderados como atributo contextual cognitivos. Então, juntamente com os atributos contextuais socioculturais, o sistema os armazena em BD provisório, utilizando como estrutura de armazenamento o *Extensible Markup Language* (XML).

Em primeiro ciclo, o sistema utiliza os termos de busca para recuperar documentos e arquivos que contenham conjuntos de palavras similares identificados anteriormente no processo de descrição. Devido à abordagem de identificação exaustiva dos termos de descrição na fase inicial, é provável que a revocação dos documentos recuperados seja alta, uma vez que termos idênticos podem ser encontrados.

A partir da recuperação da informação, o usuário deve identificar quais informações lhe parecem ser mais relevantes, selecionando conjuntos de informações que podem satisfazer sua necessidade. Do ponto de vista de sistema, um grau

numérico é atribuído de 1 a 3 para avaliação da informação recuperada, sendo 1 pouco ou nada relevante, 2 relevante e 3 completamente relevante, ou pertinente. Esta avaliação visa dar base para o sistema, então, realizar vinculações no conjunto de documentos a partir do *feedback* do usuário. A avaliação leva em conta condições subjetivas de escolha de relevância, fortemente relacionadas com a realidade específica do usuário, bem como sua história, cultura e desenvolvimento. Dessa forma, a seleção fornece para o sistema uma visão sobre os contextos linguísticos, simbólicos e interpretativos (LSI) do usuário.

Documentos considerados relevantes e documentos pertinentes, de grau 2 ou 3, terão seus termos vinculadores atualizados com os dados contextuais de perfil do usuário salvos em linguagem de marcação extensível, bem como com os termos contextuais de busca utilizados para sua recuperação. Isso propicia um refino da forma como o documento é vinculado, fazendo com que sua identificação caminhe para a especificidade, apontando de forma direta o usuário para sua necessidade. Este conjunto de marcadores contextuais que atualiza os documentos pode ser entendido como um “mapa contextual”, que leva em conta o perfil do usuário (atributos contextuais socioculturais), sua necessidade e capacidade cognitiva (atributos contextuais cognitivos) e escolha de relevância, pertinência e forma (atributos contextuais LSI).

Após diversos ciclos de busca e recuperação da informação, os documentos e arquivos deverão conter termos vinculadores mais refinados, capazes de relacionar de forma mais precisa as demandas de um usuário e o estoque informacional. A tendência do sistema, assim, é de recuperar informações com maior precisão para um usuário, tendo em vista que seus termos vinculadores terão maior grau de especificidade, a partir dos atributos contextuais. Para tanto, a busca deverá considerar não apenas os descritores ou termos de busca, mas também o perfil do usuário e especificidades das buscas apresentadas. No entanto, isso depende da forma como o usuário utiliza o sistema.

Perfis de usuário com pouca descrição e buscas alicerçadas em descritores básicos irão recuperar documentos e arquivos que contenham estas descrições básicas. No entanto, mesmo considerando esta condição, o grau de relevância e pertinência das informações recuperadas deve ser maior do que um SRI clássico,

tendo em vista a atualização constante dos termos vinculadores apresentados no presente trabalho.

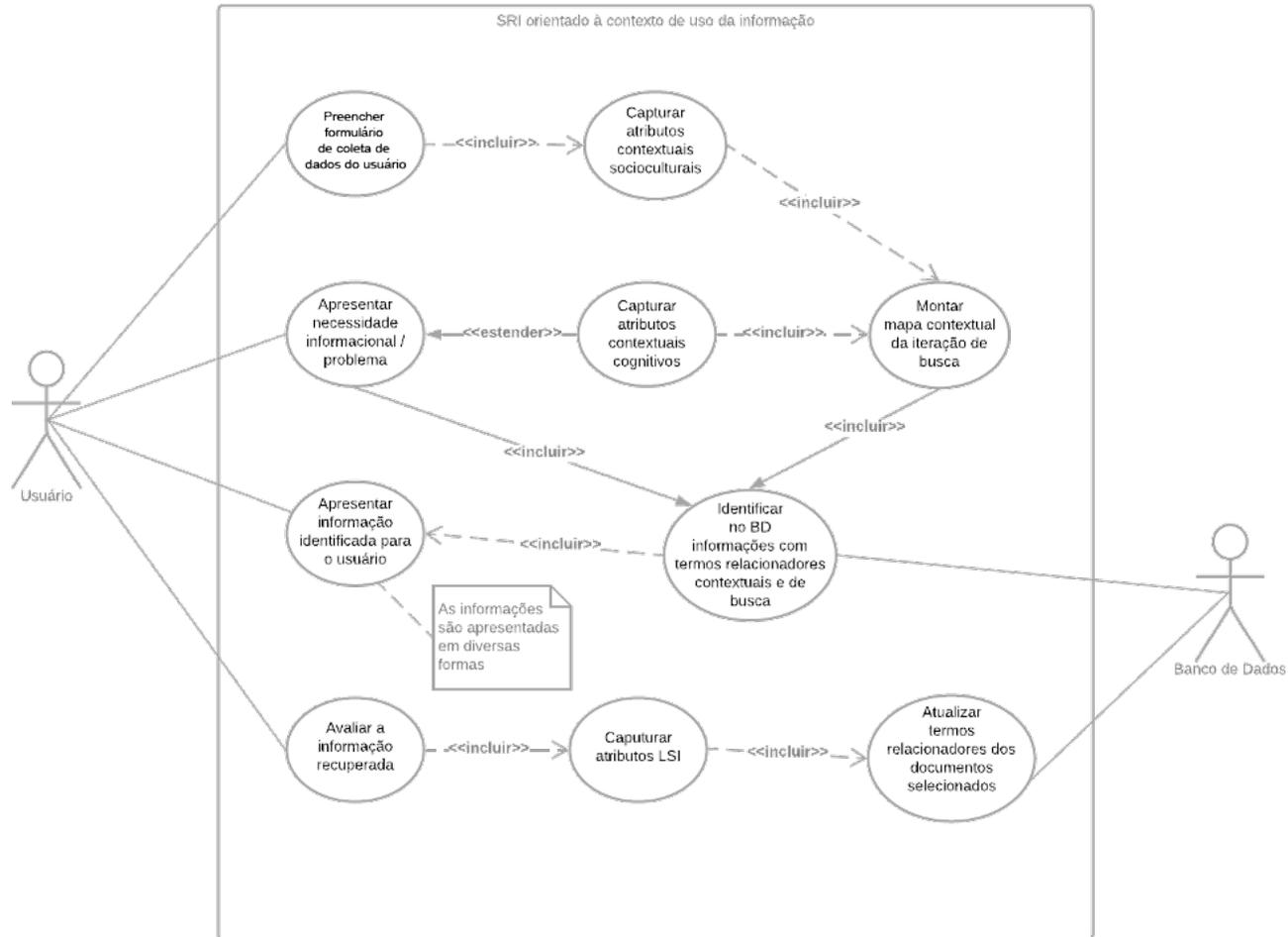
Também, após diversos ciclos de busca e recuperação da informação, o sistema deverá, a partir de aprendizado de máquina, descartar dos documentos termos vinculadores que não aparecem de forma constante para perfis de mesma característica, diminuindo assim a revocação e aumentando a precisão. Isso é importante para garantir buscas mais alinhadas com as especificidades dos usuários e evitar o fenômeno do *overfitting*, ou ajuste excessivo (SAVOLAINEN; THOMSON, 2022). Ajuste excessivo é um conceito utilizado no campo de aprendizado de máquina para descrever quando um modelo é ajustado de forma excessiva aos dados de treinamento, falhando em generalizar para novos dados. Isso acontece quando o modelo é muito complexo e acaba tentando aprender até mesmo informações insignificantes presentes nos dados de treinamento, ao invés de se concentrar apenas nas informações relevantes.

Caso o sistema mantenha estes termos vinculadores não utilizados, pode criar condições de “falsas precisões” para perfis muito particulares, reforçando a revocação do sistema para estes perfis. Também, novas informações vão ser apresentadas constantemente sem termos vinculadores treinados, a fim de garantir recuperações de informação potencialmente mais atualizadas com possibilidade de novas vinculações. Isso atende à característica de atualização constante deste tipo de sistema, conforme proposto por Ingwersen (1996).

Assim, o modelo é constantemente atualizado a depender da avaliação do usuário, melhor detalhando os termos vinculadores dos documentos e arquivos, a fim de melhor vinculá-los com os usuários, a partir de atributos contextuais de uso da informação. Entende-se, para o modelo proposto, que um maior detalhamento dos contextos dos usuários e suas necessidades, neste sentido, podem recuperar informações com maior pertinência, com documentos indexados com de forma mais especializada, em uma RI com maior precisão.

A Figura 16, em tamanho maior no Apêndice 1, apresenta a estrutura geral do modelo proposto, considerando o supracitado.

Figura 16 – Diagrama de caso de uso do modelo de SRI proposto no trabalho



Fonte: Elaborado pelo autor.

A seguir, na próxima seção, os processos são detalhados a nível de operação, explicitando suas vinculações teóricas consideradas para seu desenvolvimento.

### 4.3 Detalhamento do modelo proposto

O modelo proposto leva em consideração a natureza dos ambientes de informação digital, que são caracterizados pela produção em grande volume, alta velocidade e variabilidade, tornando impossível a representação documental prévia do que é armazenado. Além disso, não é possível definir previamente o perfil do usuário dessas informações (DAVENPORT, 2014; WHITE, 2015).

Com o auxílio de ferramentas de análise textual e imagética, como o *Knowledge Discovery in Texts* (KDT), no entanto, é possível realizar uma verificação preliminar do conteúdo, identificando palavras de alta frequência e termos específicos relacionados a substantivos, adjetivos, advérbios e outros. Esta abordagem permite identificar termos iniciais que podem ser utilizados nos primeiros ciclos de treinamento do modelo proposto.

Busca-se, de uma maneira geral, utilizar atributos contextuais de uso da informação para atualizar os termos vinculadores dos documentos e arquivos recuperados de um estoque informacional, levando em conta o *feedback* dos usuários na seleção das informações recuperadas. Isso permite caracterizar os documentos e arquivos e suas instâncias de acordo com perfis de usuários e necessidades informacionais específicas, garantindo que sejam recuperados apenas conteúdos relevantes para cada contexto.

Para tanto, são coletados atributos contextuais de uso da informação dos três campos observados no referencial teórico: sociocultural, cognitivo e linguístico/simbólico/interpretativo. Os atributos dos dois primeiros campos servem como novos qualificadores/termos vinculadores dos documentos ou arquivos recuperados, e são atribuídos a eles a partir do *feedback* do usuário, que seleciona informações recuperadas com base em relevância e pertinência. A partir daí, em caso de relevância e pertinência, os metadados capturados dos campos sociocultural e cognitivo são atribuídos aos documentos, aprimorando seus termos vinculadores.

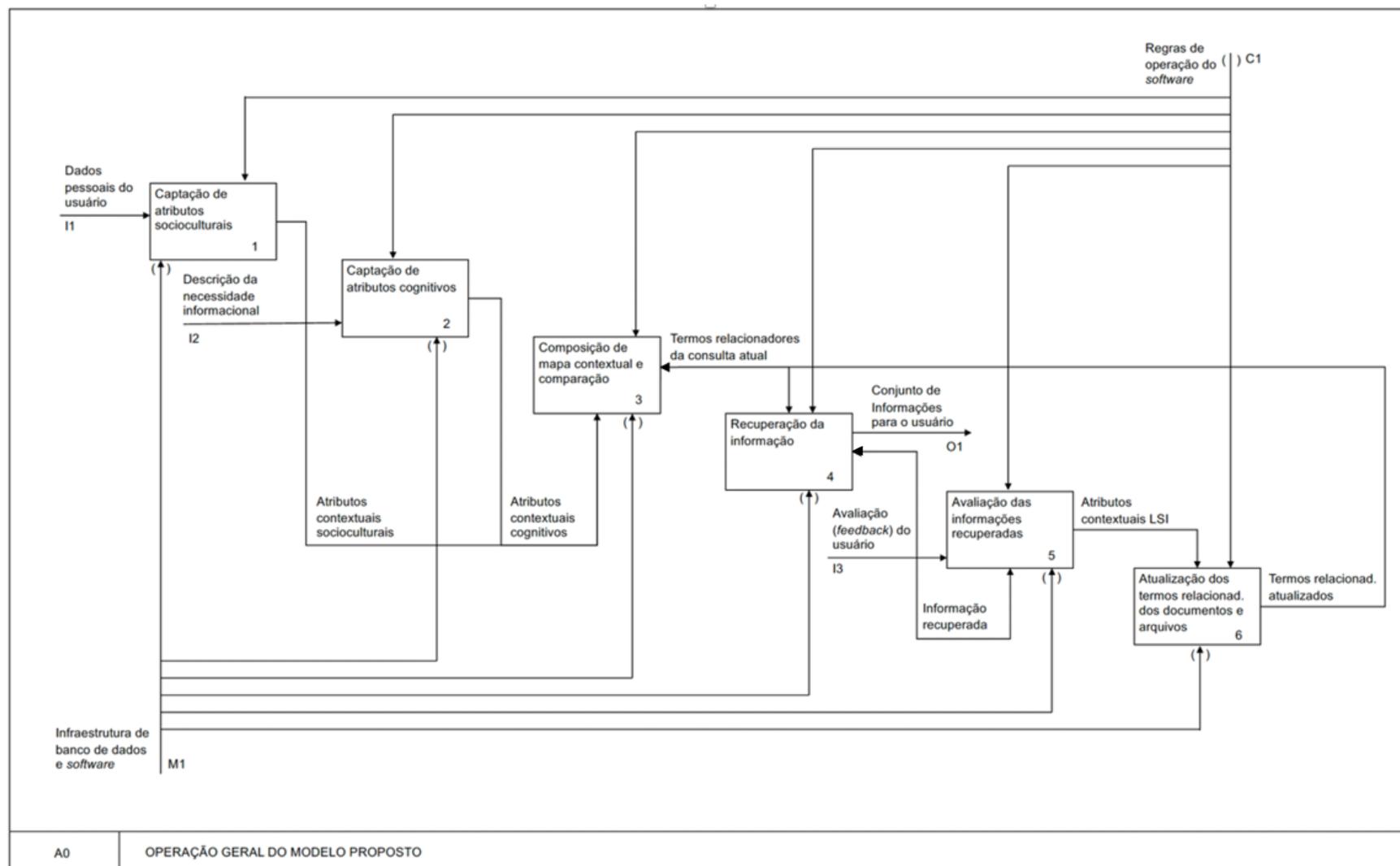
A aplicação dos novos termos vinculadores vai além da comparação com os termos de busca tradicionais, incluindo também os aspectos socioculturais e cognitivos do contexto em novas consultas. Isso garante uma conexão mais ampla não apenas entre o problema informacional e os documentos relevantes, mas também entre o perfil do usuário e os documentos que podem ser úteis. Com o tempo, ciclicamente, novos perfis de usuário vão alimentando os novos termos vinculadores dos documentos armazenados, bem como reforçando os já existentes, descartando aqueles que são menos contributivos para a qualificação do documento ou arquivo.

Tal interação cíclica possibilita a criação de novos vínculos entre necessidades informacionais, perfis de usuário e documentos estocados, permitindo recuperações mais pertinentes em consultas futuras. Isso faz com que o SRI se atualize, atendendo o proposto por Ingwersen (1999) e Ingwersen e Jäverlin (2004), sobre atualização constante do sistema, em face da atualização do perfil de usuário, social e cognitivo.

Quando uma nova consulta é feita, o sistema utiliza o mapa contextual para compará-lo com os termos vinculadores já salvos nos documentos, a fim de recuperá-los de forma específica para os usuários.

A Figura 17, em tamanho maior no Apêndice 2 apresenta este processo de forma geral, elaborado com a técnica IDEF-0.

Figura 17 - Processo de operação do modelo proposto representando em diagrama do tipo IDEF-0



Fonte: Elaborado pelo autor.

Nota-se, na representação gráfica, que os dois processos iniciais (1, 2) recebem suas entradas de operação de informações fornecidas pelos usuários, sendo elas de campo sociocultural, relativas a seu perfil, e de campo cognitivo, relativas à sua necessidade informacional e características básicas cognitivas. A partir daí, o terceiro processo (3) recebe como estrutura de mecanismo, ou ferramenta, estes atributos coletados para processar as regras de recuperação da informação. Essas regras são enviadas para o quarto processo (4), que, através de processamento de dados, recupera para o usuário informações que possam lhe ser relevantes.

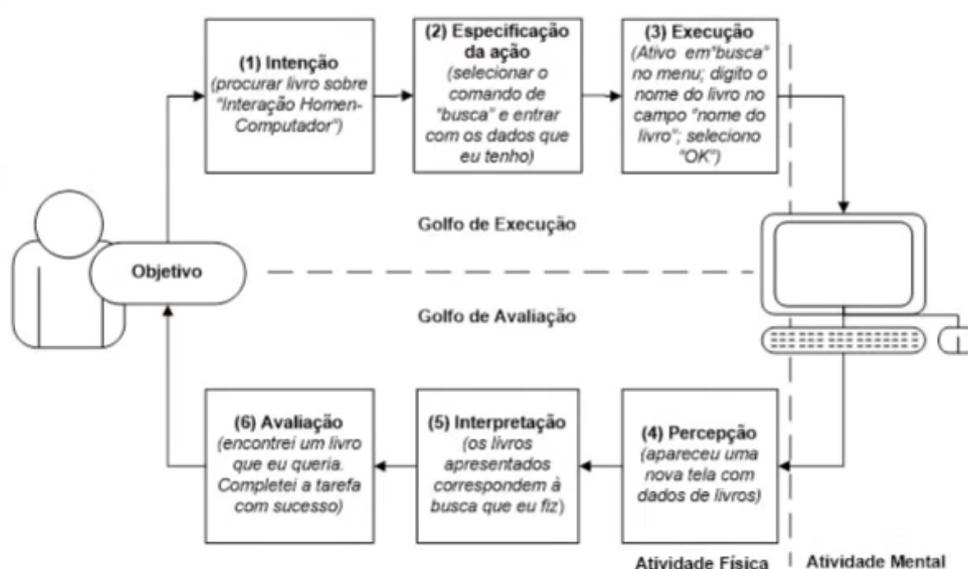
As informações recuperadas então são avaliadas no quinto processo (5), pelo usuário, que alimenta as regras do sexto processo (6) para atualização dos termos vinculadores dos documentos armazenados, vinculando contexto de uso da informação e termos vinculadores dos documentos, fortalecendo vínculos que aprimoram a busca e a recuperação da informação. Importante que todos estes processos recebem suporte de estruturas computacional de *hardware* e *software* para processamento, seguindo as regras dos algoritmos computacionais de operação de bancos de dados e dos softwares. Essas regras dizem respeito à operação do software a nível de programação.

Assim, entende-se que o contexto de uso da informação é composto por elementos que caracterizam o usuário, seu problema informacional e sua relação com a informação, tanto em termos comportamentais quanto interpretativos. Cada ciclo de busca e recuperação em um SRI ocorre de forma singular, onde a combinação destes fatores contextuais é única, gerando uma espécie de "impressão digital" da iteração de consulta. Portanto, a vinculação dos atributos socioculturais e cognitivos aos documentos e arquivos pode levar a recuperações de informação mais relevantes e pertinentes, aumentando a possibilidade de apropriação e operacionalização da informação pelo usuário.

Para desenvolver um modelo com estas características, é possível utilizar a disciplina de Interação Homem-Computador (IHC) como base. A IHC já considera o uso de contextos observados nos potenciais usuários, buscando aprimorar a comunicação entre um usuário e o computador, bem como processos desenvolvidos de interação entre usuário e sistema.

Com base nesta abordagem, os sistemas são divididos em dois golfos: execução e avaliação. Na fase de execução, são investigados os contextos que caracterizam o usuário, enquanto na fase de avaliação são determinadas as melhores formas de recuperação de informação, bem como os potenciais condições de satisfação ou sucesso no uso do sistema (BARBOSA; SILVA, 2010). Para projetar um sistema IHC, é necessário avaliar os estágios esperados em ambos os golfos, descritos por Barbosa e Silva (2010), a partir da teoria da ação de Donald Norman (1986), apresentado na Figura 18.

**Figura 18 - Teoria da ação, conforme proposto por Donald Norman (1986)**



Fonte: Barbosa; Silva (2010, p. 57) a partir de Norman (1986).

Tal processo, mesmo retirado das bases teóricas da TI, é consonante com o observado por Ingwersen (1999), Ingwersen e Jäverlin (2004) e Vickery e Vickery (2004) quando Barbosa e Silva (2010) apresentam etapa de execução e avaliação por parte do usuário. Na etapa de execução, o usuário, de forma explícita ou implícita, apresenta informações para o sistema operacionalizar seu problema, buscando dentro de seu repositório informações ou dados que possam auxiliar o usuário em sua demanda.

Tais informações são utilizadas pelo sistema de forma a alinhar a necessidade do usuário e a possibilidade de apresentação do sistema, através da especificação de

ação por parte do usuário. Após executar a ação de necessidade, ao usuário é apresentado o dado ou informação relativa à execução. O usuário, então, avalia a qualidade do recuperado, dando *feedback* para o sistema dar continuidade à um processo mais refinado, ou aprimorado em processo futuro. O sistema, neste sentido, atualiza sua política de operação com fins de aprimorar o funcionamento do sistema (BARBOSA; SILVA, 2010).

Esses dois grandes processos podem ser divididos em subprocessos relativos às ações de execução e avaliação, conforme proposição de Barbosa e Silva (2010). Para o sistema proposto pelo modelo da tese, o golfo de execução pode ser dividido em quatro subprocessos, à saber: (1) captação dos atributos contextuais socioculturais, (2) captação dos atributos contextuais cognitivos, (3) criação do mapa contextual e comparação e (4) recuperação da informação.

O primeiro subprocesso, de captação dos atributos socioculturais, compreende a identificação de atributos contextuais que qualificam o usuário da informação, relativos à sua questão situacional. Estes atributos foram explicitados no referencial teórico, observados na norma de usabilidade NBR 9241-11 (ABNT, 2021) quando essa trata do ambiente físico em que um determinado sistema opera.

São vistos, também em Talja, Heidi e Tarja (1999), que quando discutiram a questão interpretativa do contexto de um usuário, que versa sobre condicionantes contextuais do momento de mediação entre sistema e usuário, em consulta. Também em Foresti, Varvakis e Godoy Viera (2016) quando abordam a questão situacional e extralinguística de um usuário, apontando para contextos que envolvem situações de usuários subjetivas, dada sua construção como indivíduo.

Em consonância com as contribuições de Wersig e Windel (1985), que abordaram a condição situacional atual dos usuários, também em um contexto espacial-temporal, pode-se vislumbrar a importância da dimensão situacional. Nesse contexto, Choo (2003) enfatiza a relevância da questão situacional do usuário, que diz respeito ao momento em que se encontra, tanto no espaço quanto no tempo. Essa abordagem contextual espacial-temporal é explorada por Abowd e Dey (1999) ao discutirem sistemas sensíveis ao contexto da informação.

A "situação do usuário" emerge como um atributo contextual sociocultural essencial. Esse atributo está diretamente relacionado à maneira como o usuário se apresenta e interage durante a busca de informações em um SRI, como enfatizado por Dey (2001). Portanto, compreender a situação do usuário é fundamental para aprimorar a eficácia da recuperação de informações e a adaptação dos SRI às necessidades específicas de cada usuário e aos contextos em que estão inseridos.

A situação do usuário refere-se ao contexto sociocultural em que o usuário se encontra durante o processo de busca de informação em um SRI. Essa situação pode ser compreendida por meio de atributos qualificadores, como local de pesquisa, local de nascimento, idioma, profissão, renda, religiosidade, entre outros, que ajudam a identificar o indivíduo. Estes atributos podem influenciar sua compreensão social e cultural, bem como seus valores pessoais. Além disso, eles podem indicar quais tipos de conteúdo são relevantes e interessantes para o usuário durante a recuperação da informação. A questão da temporalidade também é relevante, uma vez que o momento em que uma consulta é realizada, especialmente em relação ao desenvolvimento do usuário, pode afetar a importância atribuída a uma informação recuperada.

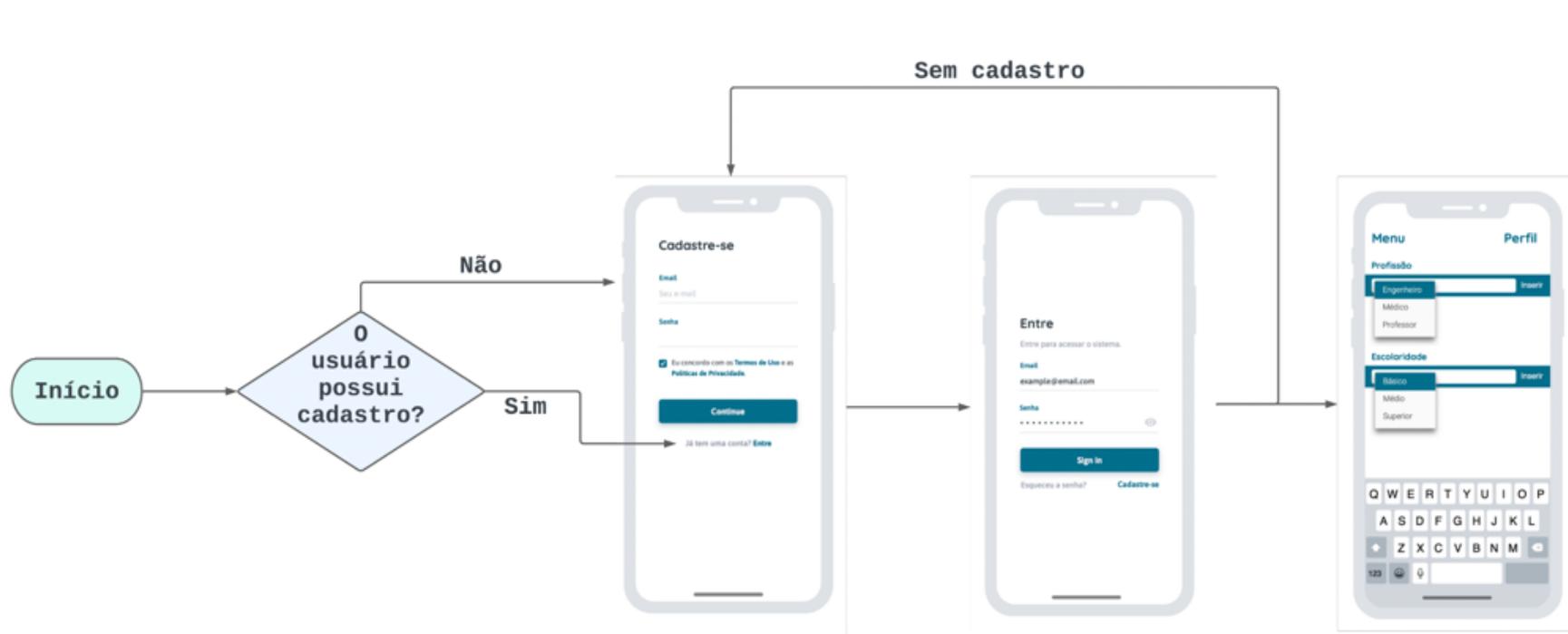
Ao utilizar este tipo de contexto, busca-se a caracterização do perfil do usuário da informação, entendendo que perfis distintos de usuário terão comportamentos informacionais distintos (WILSON, 2006). Ao adicionar essas características como termos vinculadores em documentos ou arquivos, durante uma consulta em um SRI, um usuário pode ter uma recuperação mais personalizada e vinculada à sua busca. Isso ajuda a mitigar, em parte, o problema da não identificação de usuários em ambientes informacionais digitais.

A obtenção desses atributos se dá por meio de um formulário que o usuário preenche antes de iniciar a busca ou consulta, e essas informações são posteriormente armazenadas em seu perfil dentro do sistema. O formulário solicita informações situacionais que podem enriquecer a caracterização do usuário. É relevante destacar que determinados atributos, como a dimensão espacial e temporal, são capturados de maneira automática. Tal processo de coleta por meio do formulário é essencial, uma vez que esses atributos são de difícil acesso sem uma comunicação falada ou escrita, não sendo explicitamente evidentes.

Quando o usuário acessa o sistema, tem a opção de indicar se já possui uma conta para realizar consultas. Caso afirmativo, é direcionado para a tela de preenchimento de atributos socioculturais, onde pode atualizar ou preencher seus dados. Se o usuário não possui uma conta, será solicitada a criação de uma para obter acesso. Assim que a conta for criada e liberada, o usuário poderá acessar o sistema.

A tela que representa este processo é apresentada na Figura 19, indicando a forma como o usuário será abordado pelo sistema.

Figura 19 - Interface inicial de acesso do usuário ao SRI do modelo proposto



Fonte: Elaborado pelo autor.

Como visto, além da captação direta das informações dos usuários, dados espacial-temporais também são captados de forma automática. Neste contexto, as informações obtidas podem ser armazenadas em um arquivo com a linguagem de marcação extensível. Esta linguagem foi desenvolvida para permitir que informações de documentos sejam organizadas de forma hierárquica, utilizando elementos ou *tags* que representam cada componente do que se está armazenando. Isso possibilita uma análise posterior mais detalhada das informações salvas, pois as organiza em forma de metadados de fácil leitura por técnicas como descoberta de conhecimento de dados e em texto (SILVA, 2020). A partir daí é possível a recuperação e organização desses dados em termos vinculadores, que são utilizados para compor o “mapa contextual” para atualização dos documentos.

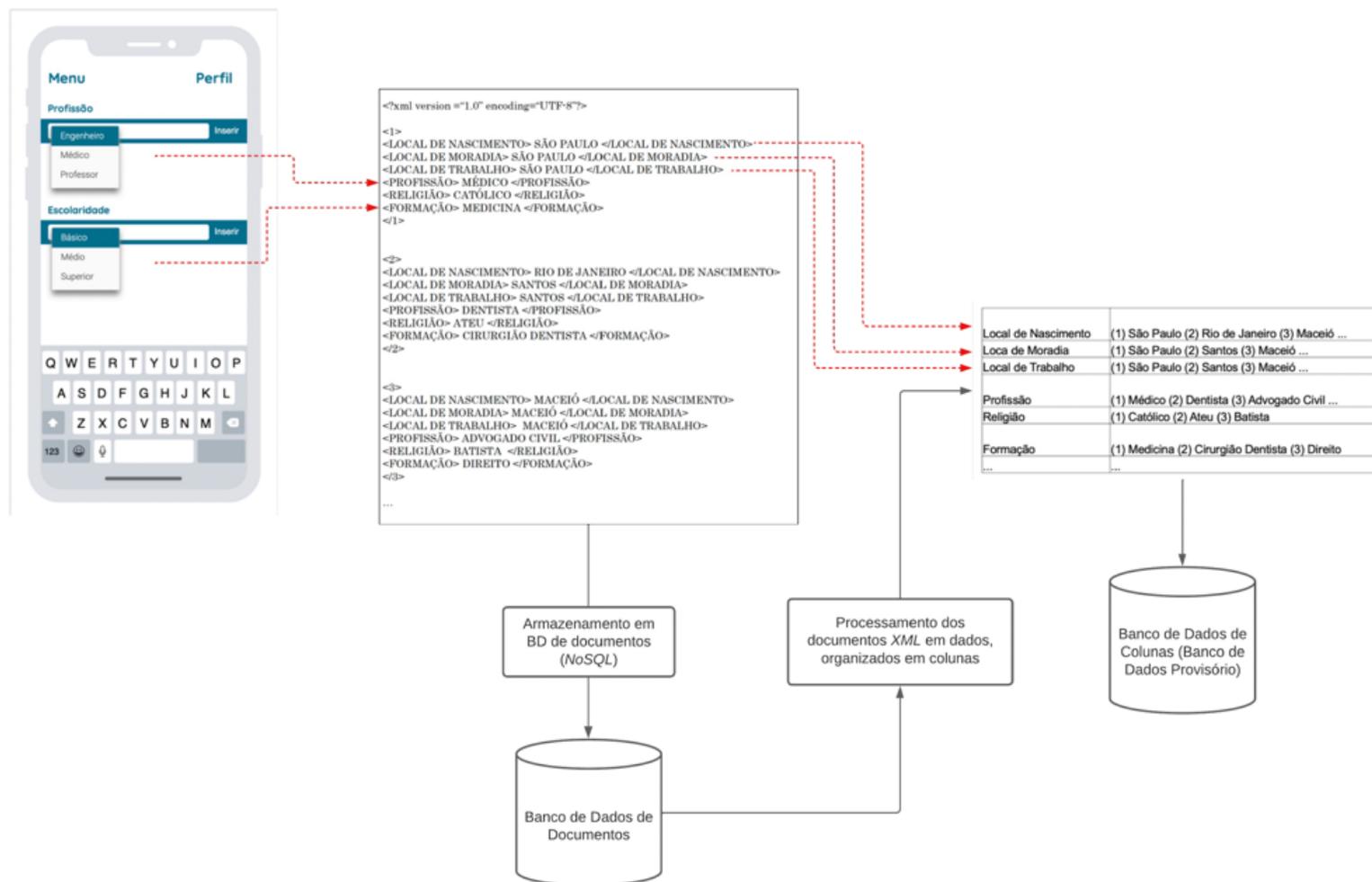
Com o objetivo de facilitar a organização dos dados coletados, a estrutura de armazenamento do estoque informacional será feita por meio de um BD NoSQL. Os bancos de dados NoSQL são não relacionais e foram desenvolvidos para lidar com grandes volumes de dados, com mais flexibilidade do que os bancos de dados relacionais tradicionais, conhecidos como SQL. Existem duas formas de armazenamento para essas informações: (1) bancos de dados de documentos, para armazenar arquivos XML com dados retirados dos formulários; e (2) bancos de dados por colunas, para organizar de forma mais rápida os atributos extraídos do processo de descoberta de conhecimento em textos, criando termos vinculadores.

No preenchimento do formulário de características socioculturais, os dados são coletados e utilizados para criar um documento de texto no formato XML. O XML é uma linguagem de marcação que organiza os dados hierarquicamente utilizando marcadores personalizadas para definir elementos e atributos. Este documento é então armazenado em um BD específico para processamento interno de dados em arquivos de texto.

Em seguida, o documento passa por um processo de descoberta de conhecimento em textos, onde palavras-chaves retiradas do texto, a partir de regras prévias, são coletadas e organizadas em um BD colunar.

No BD colunar, cada linha representa um campo do formulário, e cada preenchimento do formulário corresponde a um índice numérico nesta linha. Isso facilita o processo de recuperação de informações neste tipo de BD. A Figura 20 explicita este processo.

Figura 20 - Captação de atributos socioculturais a partir de formulário de preenchimento do usuário



Fonte: Elaborado pelo autor.

O referido procedimento é necessário porque parte do preenchimento do formulário de coleta de atributos socioculturais envolve o uso de descritores diretos, que consistem em um conjunto limitado de palavras, bem como o preenchimento dissertativo em linguagem natural. A coleta por meio de descritores fornece informações sobre regionalismos, profissões, religiosidade, renda e outros aspectos diretamente caracterizadores. O campo de descrição dissertativa permite que o usuário forneça uma descrição mais detalhada do seu perfil, usando linguagem natural. As regras para essa captação são prévias, e devem considerar substantivos e adjetivos relativos aos campos, considerados no treino do KDT.

Após salvar estas informações em um arquivo XML, é possível processar o texto escrito em linguagem natural utilizando-se a técnica de KDT. A técnica KDT utilizada na tese consiste em três etapas principais. Primeiramente, realiza-se o pré-processamento de texto, que envolve a limpeza e preparação do texto para análise, incluindo remoção de pontuação, *stop words*, lematização ou *stemming*<sup>20</sup>, e outras técnicas de normalização. Em seguida, na etapa de extração de recursos e análise, são identificadas características relevantes do texto, como palavras-chave, entidades nomeadas, relações entre entidades, tópicos, temas principais e sentimentos expressos.

Além disso, são identificados padrões, tendências e relações nos dados textuais, utilizando técnicas como aprendizado de máquina e mineração de texto. Estas identificações são realizadas por meio da comparação da base de análise com uma base já existente, verificando os graus de similaridade entre o texto em questão e o que já foi pré-treinado e armazenado na base de comparação. Por fim, os dados são organizados e armazenados para uso futuro, para a criação de mapas contextuais que funcionarão como termos vinculadores para os documentos.

---

<sup>20</sup> “*Stop words* são palavras comuns que não importam muito para a análise de texto, como “e”, “o” e “para”. Elas são tiradas fora durante o pré-processamento para focar nas palavras importantes. Lemmatização é quando palavras são reduzidas à sua forma básica ou raiz, como “correndo” para “correr”. Isso ajuda a tratar diferentes formas da mesma palavra como uma só, facilitando a análise. Stemming é um processo mais simples de reduzir palavras à sua forma raiz, removendo sufixos e prefixos, como “correndo” para “corr”. Isso também ajuda a tratar diferentes formas da mesma palavra como uma só, mas é mais rápido e menos preciso que a lematização” (GÉRON, 2019).

As regras de definição do que deverá ser identificado e salvo (palavras, expressões e termos) devem considerar, de forma exaustiva, todas as potenciais palavras qualificadoras de um usuário, sendo elas adjetivos, substantivos, advérbios e outros elementos linguísticos descritores. Estas palavras e termos são então coletados e armazenados em um BD colunar, organizadas por índice, conforme mostrado na Figura 21.

Um BD colunar NoSQL é um tipo de BD não relacional que armazena os dados em formato colunar, em vez de armazená-los em linhas como em bancos de dados relacionais tradicionais. Em um BD colunar, os dados de cada coluna são armazenados de forma contígua em disco, o que proporciona benefícios significativos em termos de desempenho e eficiência de armazenamento (WHITE, 2015).

Um BD colunar funciona armazenando os dados em colunas, em vez de armazená-los em linhas como em bancos de dados tradicionais. Cada coluna contém os valores correspondentes para essa propriedade específica em todas as linhas. Isso permite um armazenamento mais eficiente e um processamento mais rápido das consultas.

Esta abordagem difere de um modelo tradicional relacional, adotando um modelo de colunas indexado. Isso dá mais agilidade ao processo de recuperação de dados neste tipo de banco, necessária para a estrutura de mapeamento do modelo proposto.

**Figura 21 – Modelo de organização de dados em um banco de dados NoSQL do tipo colunar**

CAMPO DE INTERESSE	ÍNDICE	DADO
Local de Nascimento	(1) São Paulo	(2) Rio de Janeiro (3) Maceió ...
Cidade de residência	(1) São Paulo	(2) Santos (3) Maceió ...
Bairro	(1) Moema	(2) Gonzaga (3) Farol ...
Cidade em que exerce atividade profissional	(1) São Paulo	(2) Santos (3) Maceió ...
Ocupação	(1) Médico	(2) Dentista (3) Advogado de direito imobiliário ...
Religião	(1) Católico	(2) Ateu (3) Batista
Formação	(1) Medicina	(2) Cirurgião Dentista (3) Direito
Posição GPS no momento da consulta	(1) Latitude: -22.9068, Longitude: -43.1729	(2) Latitude: 37.7749, Longitude: -122.4194 (3) Latitude: -33.8679, Longitude: 151.2073
...	...	...

Fonte: Elaborado pelo autor.

O BD colunar, assim, se torna repositório provisório destes dados, que os recebe e os organiza de forma a criar o “mapa contextual” inicial, relativo àquele perfil de usuário.

Após o preenchimento do formulário de caracterização de atributos contextuais socioculturais por meio do formulário de captação, o usuário é redirecionado para a tela de consulta. Antes de descrever este processo, é importante destacar duas observações: o preenchimento dos atributos socioculturais é realizado no perfil do usuário, visando aprimorar os resultados de recuperação de informação para aquele indivíduo específico. No entanto, as informações coletadas são organizadas e utilizadas para o treinamento geral do sistema. Isso significa que as informações socioculturais de um usuário em particular são usadas para aprimorar o processo de recuperação de informação para todos os outros usuários deste sistema.

Também, este processo não é estático e requer atualização contínua. À medida que os usuários atualizam suas características ao longo do tempo, influenciados por diversos fatores, como interações com outros indivíduos, crenças e valores (WEBER, 1977), é importante considerar a atualização dos dados coletados para garantir uma

correspondência mais precisa entre a consulta e as informações recuperadas. Isso envolve o aperfeiçoamento destes atributos que serão utilizados como índices de busca. Assim, quando o sistema percebe que a seleção das informações recuperadas apresenta menos da metade das considerações relevantes, solicita ao usuário a atualização de seus atributos, a fim de melhorar a vinculação entre a busca e a recuperação de informações. No entanto, a decisão de atualização cabe ao próprio usuário.

Isso posto, o usuário é direcionado para a tela de consulta, onde é solicitado que ele descreva sua necessidade ou problema, visando recuperar informações relevantes e pertinentes que possam capacitá-lo a agir. Esta descrição é realizada em linguagem natural, incentivando o usuário a fornecer o máximo de detalhes possível sobre seu problema, da maneira que considerar mais abrangente e completa. Busca-se, nessa etapa, a captação de atributos cognitivos do usuário, relacionados a seu estado anômalo de conhecimento, conforme explicitado por Belkin (1980), seu idioma, expressões e regionalismos, gírias, jargões entre outros, que indiquem o grau de conhecimento de um usuário sobre um determinado tema ou assunto.

De acordo com Foresti, Varvakis e Godoy Viera (2016) e Choo (2003), o contexto extralinguístico e cognitivo pode fornecer informações sobre o nível de conhecimento de um usuário em relação a um determinado assunto. Estas informações podem ajudar a identificar quais relações prévias o usuário tem com a informação que busca, a partir da maneira como formula suas perguntas, o que pode indicar parte de seu comportamento em relação à informação necessária (GUINCHAT; MENO, 1994).

Em um SRI, é responsabilidade do usuário formular uma expressão de busca que descreva adequadamente sua necessidade de informação (ARAÚJO JÚNIOR, 2005). Ao usar certas expressões em uma consulta, um usuário pode indicar maior ou menor familiaridade com o assunto pesquisado. Além disso, ao considerar um idioma específico de consulta, o usuário pode acabar limitando sua busca ao não usar termos compreensíveis em outras línguas para o mesmo tópico.

É importante destacar que a maior dificuldade enfrentada pelos usuários é saber quais termos usar para encontrar documentos que atendam às suas necessidades específicas (FERNEDA, 2003). Esta dificuldade, quando trabalhada

como atributo de contextualização cognitiva, pode ser usada como um facilitador de indexação entre formas e densidades de informação e perfis com maior ou menor familiaridade com ela.

Neste sentido, a captação de atributos cognitivos, ou seja, atributos relativos aos termos de consulta utilizados, bem como idioma, regionalismos, jargões e afins, envolve os processos de (1) identificação de termos específicos de busca que indiquem a necessidade informacional do usuário e (2) a identificação de expressões que levem a indicar a profundidade de conhecimento de um usuário sobre um determinado tema ou tópico. Tecnicamente, ambos processos utilizam as mesmas ferramentas da captação de atributos socioculturais. As informações são armazenadas inicialmente em documento do tipo XML para posterior mineração e organização dos dados em BD do tipo colunar. Isso pois as informações apresentadas pelo usuário para a realização de da busca por informações são feitas em linguagem natural, o que demanda de trato para identificação de marcadores individuais.

Ao consultar um BD colunar, ao invés de recuperar todas as colunas para cada linha, é possível selecionar apenas as colunas relevantes para a consulta. Isso reduz a quantidade de dados que precisam ser lidos e transferidos, resultando em consultas mais rápidas (WHITE, 2015).

A identificação dos termos de busca considera as palavras, conjuntos de palavras e termos específicos que possam vincular a consulta do usuário a um documento ou arquivo que contenha informações relevantes para o usuário. Durante este processo, busca-se selecionar termos relevantes para a consulta, a fim de identificar abrangentemente, e de forma exaustiva, todos os termos que possam compor a expressão de busca.

Isso é feito a partir de mineração de palavras e conjuntos de palavras retirados do arquivo XML composto pela expressão de busca do usuário, em linguagem natural. As regras consideradas para este processo de mineração devem envolver também adjetivos, substantivos, advérbios e outros elementos linguísticos descritores do que está contido no texto. Como identificar uma temática é um processo complexo, atenta-se para a seleção de palavras e termos com o maior número de repetições dentro do texto, sem qualquer julgamento de valor. Estas repetições podem dar lastro básico para vinculação de consulta e recuperação, para posterior refino. O processo de

extração também considera treino anterior da ferramenta de descoberta de conhecimento em textos para identificação destes termos para seleção e armazenamento, conforme explicitado no processo de captação de atributos socioculturais.

Após a identificação dos termos de busca, durante o processo de captação e geração de arquivo XML, também são levantadas expressões idiomáticas, regionalismos, jargões, gírias e outros elementos semelhantes. Estes elementos são selecionados e armazenados como qualificadores em repositório colunar, utilizando a técnica de KDT apresentada anteriormente. No processo de descoberta de conhecimento em textos, após a expressão por parte do usuário do seu problema informacional, o texto gerado é também armazenado em um arquivo XML. Este arquivo é salvo em um BD de documentos para processamento.

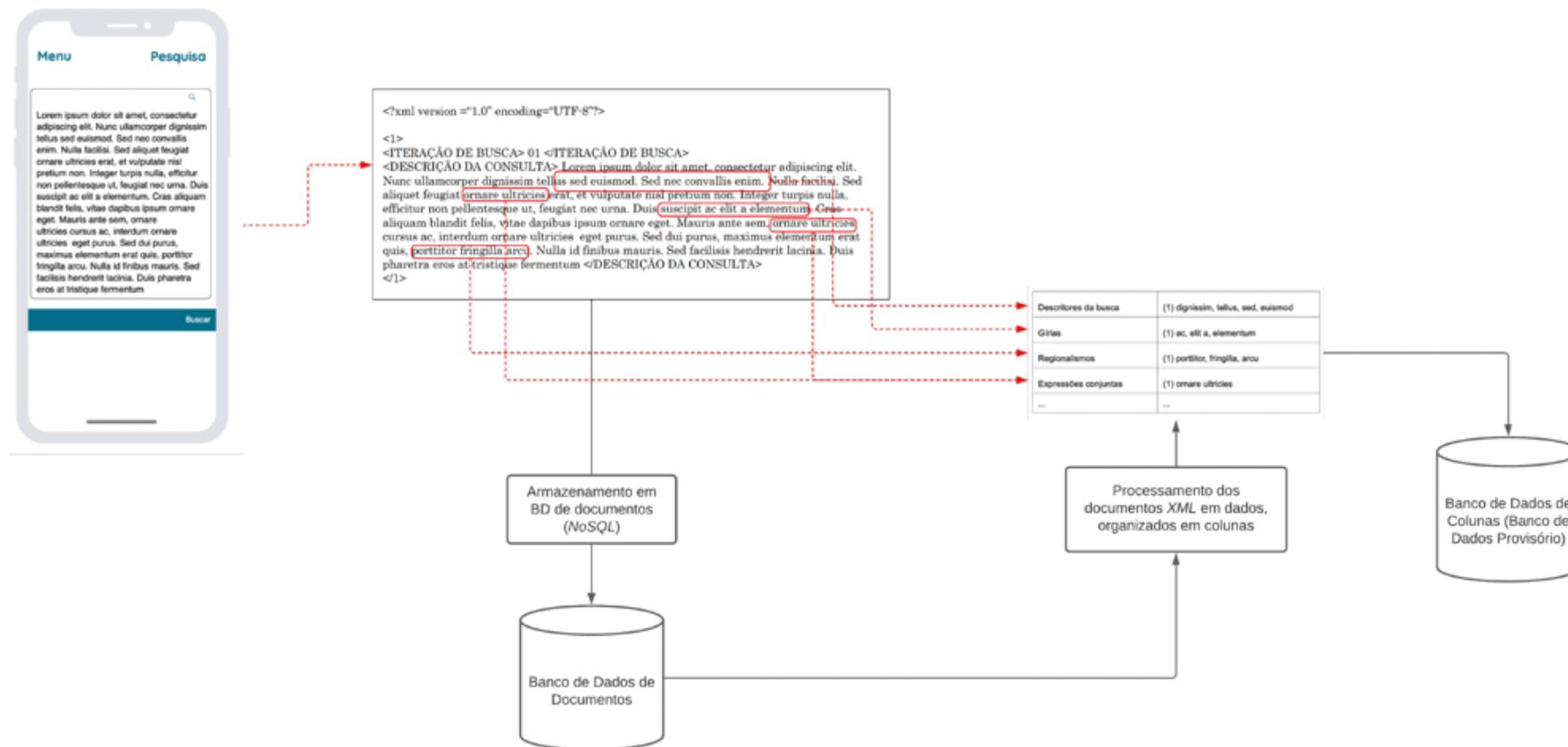
Um BD de documentos NoSQL é um tipo de BD não relacional projetado para armazenar, recuperar e gerenciar dados em formato de documentos. Ao contrário dos bancos de dados relacionais tradicionais, que são baseados em tabelas e esquemas fixos, os bancos de dados de documentos NoSQL permitem que os dados sejam armazenados em documentos flexíveis, geralmente no formato *JavaScript Object Notation* (JSON), *Binary JSON* (BSON) ou XML (WHITE, 2015).

Neste tipo de BD, os documentos são as unidades básicas de armazenamento e organização dos dados. Cada documento é uma estrutura autocontida que pode conter qualquer número de pares chave-valor, semelhante a um objeto em programação orientada a objetos. Estes documentos são armazenados em coleções, que são agrupamentos lógicos de documentos.

O processo de descoberta de conhecimento em textos, orientado a regras de identificação de substantivos, adjetivos, termos repetidos e afins que possam identificar temáticas específicas no texto, bem como gírias, jargões e regionalismos, identifica essas palavras no documento XML e as salva em BD colunar, criando neles os índices apresentados anteriormente. O processo é similar ao de captação de atributos socioculturais, e essencial para a geração do mapa contextual.

A Figura 22 explicita este processo inicial.

**Figura 22 - Processo de KDT para captura de atributos contextuais socioculturais e cognitivos do modelo proposto no trabalho**



Fonte: Elaborado pelo autor

Tanto os termos de busca quanto os descritores da expressão de busca podem representar o uso contextual da informação por um usuário, especialmente quando consideramos seus aspectos socioculturais. De acordo com Dey (2000), o contexto em que um usuário se encontra é único para cada busca e recuperação de informação. Isso significa que as particularidades contextuais devem ser consideradas de maneira específica em cada busca, permitindo assim a qualificação de uma combinação de consulta e recuperação de informação com base em seus atributos contextuais (DEY, 2001).

Entende-se que o usuário empreende sua busca por informação a depender de uma necessidade específica (FIGUEIREDO, 1994) e que essa busca acontece em um contexto específico (ABOWD; DEY, 1999). Dessa forma, entende-se existir potencial cruzamento entre problema informacional e o contexto em que o usuário se encontra, no momento da consulta ao SRI. Em outras palavras, entende-se que determinados perfis de usuário terão certos tipos de necessidades informacionais, em contextos potencialmente similares. A combinação assim, de atributos contextuais socioculturais e cognitivos, conforme observado em Talja, Heidi e Tarja (1999) , Foresti, Varvakis e Godoy Viera (2016), Choo (2003) e Wersig e Windel (1985), e a necessidade informacional do usuário, pode criar um “mapa contextual” capaz de qualificar uma instância de busca e recuperação da informação.

Isso posto, o terceiro subprocesso considerado no fluxo de execução do modelo é a criação do mapa contextual e comparação. A criação do mapa contextual envolve a combinação dos dados contextuais socioculturais e cognitivos extraídos dos documentos XML, preenchidos por meio do formulário inicial e da tela de consulta. Estes dados contextuais, como visto, são obtidos destes documentos por meio de processo de descoberta de conhecimento em textos e são armazenados temporariamente em um BD colunar. Este armazenamento visa uma futura avaliação de elegibilidade de uso, para posterior vinculação dos documentos avaliados pelos usuários na recuperação da informação. Apesar de listado como terceiro processo, este processo não ocorre de forma subsequente aos dois anteriores, mas de forma paralela.

Ao preencher tanto o formulário quanto o campo de busca, o documento XML correspondente é gerado e armazenado em um BD documental. O processo de KDT

é iniciado automaticamente para cada documento, capturando os termos relevantes durante o processamento. Assim, cada consulta gera uma nova chave de índice no BD colunar, contendo os dados capturados para compor o mapa contextual daquela consulta. Cada índice desta base de dados colunar pode ser entendido como mapa contextual. A Figura 23 representa este mapa contextual.

**Figura 23 - Mapa contextual composto por atributos socioculturais e cognitivos capturados em processo de KDT no modelo proposto no trabalho**

**MAPA CONTEXTUAL**

<b>ATRIBUTOS CONTEXTUAIS SOCIOCULTURAIS</b>	
Local de Nascimento	(1) São Paulo (2) Rio de Janeiro (3) Ma
Loca de Moradia	(1) São Paulo (2) Santos (3) Maceió ...
Local de Trabalho	(1) São Paulo (2) Santos (3) Maceió ...
Profissão	(1) Médico (2) Dentista (3) Advogado Civil ...
Religião	(1) Católico (2) Ateu (3) Batista
Formação	(1) Medicina. (2) Cirurgião Denstista (3) Direito
...	...
<b>ATRIBUTOS CONTEXTUAIS COGNITIVOS</b>	
Descritores da busca	(1) dignissim, tellus, sed, euismod
Gírias	(1) ac, elit a, elementum
Regionalismos	(1) porttitor, fringilla, arcu
Expressões conjuntas	(1) ornare ultricies
...	...

Fonte: Elaborado pelo autor.

A partir daqui novos dois subprocessos ocorrem (1) comparação com os termos vinculadores dos documentos já armazenados no repositório para recuperação à um usuário e (2) espera do processo de avaliação documental, a fim de eventualmente vincular atributos aos documentos, como termos vinculadores.

Para recuperar informações para um usuário, os dados do índice armazenados no BD colunar, gerados através da captação de KDT, são comparados com os índices

dos documentos armazenados no repositório digital de informações. Os índices dos documentos armazenados no repositório digital são gerados a partir dos metadados da análise inicial dos documentos, ou por atualização do sistema, utilizando-se a técnica proposta no presente trabalho.

Importante indicar que os termos vinculadores dos documentos armazenados não estão alocados juntos dos documentos, mas sim em BD colunar que correlaciona uma chave de índice de dados para aqueles documentos e seus termos vinculadores. Isso é feito para aprimorar o processo de RI, sem precisar de varredura total em todos os documentos.

A comparação de índices tem como objetivo promover a recuperação da informação por representação. Essa comparação é realizada através do sistema, que compara os conjuntos de dados contextuais armazenados nos termos vinculadores documentais e nos mapas contextuais. Em caso de comparação positiva, em qualquer nível, são associados documentos e consultas para recuperação. Espera-se, nas primeiras interações de busca e RI uma alta taxa de revocação, tendo em vista que poucos termos são considerados para comparação e vinculação entre termos vinculadores de documentos e mapas contextuais.

O sistema realiza uma comparação intensiva entre termos vinculadores e mapas contextuais, percorrendo todos os termos vinculadores dos bancos de dados que os contenham. A comparação ocorre entre o mapa contextual da consulta e os termos vinculadores armazenados em um BD colunar. Para garantir a eficiência do processo, os termos vinculadores não são diretamente atribuídos aos documentos, mas sim associados por meio de um índice no BD colunar. Essa abordagem visa agilizar o processo de comparação.

A Figura 24 apresenta a forma de comparação entre mapa contextual e termo vinculador de documentos.

**Figura 24 - Comparação ente termos vinculadores dos documentos e mapa contextual gerado pela captura de atributos de contexto de uso socioculturais e cognitivos**



Fonte: Elaborado pelo autor.

Para verificar possíveis semelhanças entre elementos armazenados nos termos vinculadores de documentos e nos índices de mapas contextuais coletados, utiliza-se a tecnologia GPT. O GPT é capaz de identificar similaridades mesmo em dados incompletos por meio de sua rede neural do tipo *Transformer*, que examina os pontos de maior atenção ou repetição nos nodos de entrada (VASWANI *et al.*, 2017).

O GPT, por meio de *embeddings* contextuais, emprega vetores para a comparação de textos. Cada palavra ou segmento textual é associado a um vetor, uma representação numérica que se assemelha a uma seta em um espaço multidimensional (NEELAKANTAN *et al.*, 2022). Nesse contexto, duas abordagens primárias de comparação se destacam: a similaridade de cosseno, que avalia a proximidade entre os vetores, por meio do ângulo formado entre eles, e a distância Euclidiana, que quantifica a extensão direta entre esses pontos no espaço vetorial. A similaridade de cosseno reflete-se em ângulos menores indicando maior similaridade entre os textos, enquanto a distância Euclidiana traduz-se em menor separação entre os vetores, denotando maior semelhança (VASWANI *et al.*, 2017; NEELAKANTAN *et al.*, 2022).

Essa metodologia possibilita ao GPT identificar padrões mesmo em dados parcialmente preenchidos, caracterizados por uma alta taxa de repetição. Quando tal situação ocorre, o sistema é orientado a recuperar documentos, arquivos ou casos com termos correlatos para avaliação pelo usuário. Tal direcionamento reveste-se de significativa relevância, pois, ao confrontar índices de documentos por meio desses mapas contextuais, os termos disponíveis inicialmente são aqueles previamente identificados, visando incorporar uma representação inicial abrangente e detalhada (NEELAKANTAN *et al.*, 2022).

Esta tecnologia é escolhida para este processo porque é mais rápida do que os métodos de comparação tradicionais, que exigem um processo de varredura total do sistema, demandando de atributos idênticos para paridade. Com o GPT, é possível realizar a varredura dos índices do BD colunar, identificando de forma parcial similaridade para efetuar, assim a recuperação da informação. Além disso, ela lida melhor com repetições de palavras e conjuntos de termos, não se limitando apenas a conjuntos de palavras previamente treinados, conforme apontado por Vaswani *et al.* (2017).

Durante o processo de comparação, os dados obtidos como atributos contextuais são mantidos no BD colunar, aguardando a avaliação do usuário. Se o usuário considerar as informações relevantes no conjunto de arquivos e documentos recuperados, o sistema tem a capacidade de reindexar estes documentos, criando índices com base nos atributos contextuais previamente coletados.

O processo de criação do mapa contextual e comparação com lista de termos vinculadores dos documentos armazenados leva ao quarto e último subprocesso do golfo de execução, que é o processo de recuperação da informação. O processo de recuperação da informação é feito através da apresentação dos documentos e arquivos com termos vinculadores em paridade com os mapas contextuais. Importante indicar que, para o modelo proposto, não são recuperados apenas documentos e arquivos, mas também, são compostas e apresentadas novas formas de informação para aprimorar a acessibilidade da informação, bem como sua potencialidade de relevância e apropriação por um usuário.

Isso se alinha com o apresentado por Araújo (1995), Ingwersen (1996) e Ingwersen e Jäverlin (2004) que indicaram que SRIs e SRIs digitais, no geral, se assentam na prática de recuperação documental, focando exclusivamente no conteúdo, sem se atentar para a forma da informação.

Além disso, o sistema proposto no modelo incorpora novas informações que são recentemente indexadas de forma exaustiva, sem considerar o processo de refinamento dos atributos contextuais. Esta abordagem é adotada para evitar o problema do ajuste excessivo, conforme apontado por enunciado por Ingwersen e Jäverlin (2005), onde o sistema fica limitado a recuperar os mesmos conjuntos de dados, sem a possibilidade de obter novas informações (SAVOLAINEN; THOMSON, 2022).

Dessa forma, o modelo apresenta três tipos de recuperação da informação: (1) documental, (2) adaptada e (3) recente. A RI documental recupera para o usuário, a partir de método computacional básico os documentos e arquivos potencialmente relevantes para um usuário, de forma que ele os avalie.

Ademais dessa recuperação, estes mesmos arquivos e documentos são apresentados para o usuário em formas diversas, com diferentes níveis de adensamento de conteúdo. Isso é feito para atender o enunciado por Ingwersen e

Jäverlin (2005), quando indicam os SRIs deveriam ser projetados considerando-se a relevância para seu usuário a nível também de acessibilidade, ou forma. Isso faz com que o usuário tenha maior poder de apropriação da informação, melhorando as chances de conseguir operacionalizar com a informação recuperada.

O processo de geração de texto também pode ser realizado usando a tecnologia GPT de forma generativa. Com treinamento prévio, é possível fazer com que essa tecnologia adapte os textos em diferentes níveis de explicação, variando do mais simples ao mais complexo. No caso deste modelo, são adotados três níveis de explicação: simplificado, intermediário e complexo. Estes níveis têm como objetivo expressar a informação de maneira simples e analógica, de acordo com o contexto do usuário, utilizando exemplos aplicados para facilitar a compreensão, e, quando necessário, oferecendo uma explicação detalhada com o uso de jargões e termos técnicos.

Também, de forma generativa, o modelo pode adaptar as informações em infográficos e imagens, geradas a partir do texto para auxílio no seu entendimento, a partir das tecnologias GPT, que modelam, a partir de treinos anteriores, imagens em função de textos e termos de busca. Utilizando como base os documentos recuperados, imagens explicativas dos mesmos, bem como reestruturações de adensamento textual, podem prover novas formas de apresentação da informação, atendendo o explicitado por Ingwersen e Jäverlin (2005) e Araújo (1995). O Quadro 5 explicita os níveis possíveis de apresentação da informação, com o esperado, conforme proposto pelo modelo.

**Quadro 5 - Forma de apresentação da informação recuperada no modelo proposto**

<b>Nível de Explicação</b>	<b>Descrição</b>	<b>Exemplo de Texto</b>	<b>Forma de Apresentação</b>
Simplificado	Expressa a informação de maneira simples e analógica, utilizando exemplos aplicados para facilitar a compreensão.	"A tecnologia GPT é como um assistente de escrita que gera texto com base no que foi aprendido anteriormente."	Texto explicativo
	Expressa a informação de maneira simples e analógica, utilizando exemplos aplicados para facilitar a compreensão.	-	Infográfico mostrando o funcionamento do processo de geração de texto com a tecnologia GPT
Intermediário	Oferece uma explicação mais detalhada, utilizando jargões e termos técnicos quando necessário.	"A tecnologia GPT utiliza modelos de linguagem treinados em grandes conjuntos de dados para gerar texto de forma generativa, adaptando-o em diferentes níveis de explicação."	Texto detalhado com jargões e termos técnicos
	Oferece uma explicação mais detalhada, utilizando jargões e termos técnicos quando necessário.	-	Imagem mostrando a estrutura de uma rede neural utilizada pela tecnologia GPT

continua...

conclusão...

**Quadro 5 - Forma de apresentação da informação recuperada no modelo proposto**

Complexo	Fornece uma explicação detalhada com o uso de jargões e termos técnicos, proporcionando um nível mais avançado de compreensão.	"A tecnologia GPT emprega técnicas de aprendizado profundo, como redes neurais, para realizar o processo de geração de texto generativo, possibilitando a adaptação em múltiplos níveis de abstração e complexidade."	Texto detalhado com jargões e termos técnicos
	Fornece uma explicação detalhada com o uso de jargões e termos técnicos, proporcionando um nível mais avançado de compreensão.	-	Imagem mostrando uma representação visual complexa dos processos envolvidos na geração de texto com a tecnologia GPT

Fonte: Elaborado pelo autor.

A recuperação de novos documentos é realizada considerando as entradas recentes no repositório de dados, que possui características de ambientes informacionais digitais, e comparando-as com os documentos já armazenados. São considerados para recuperação e avaliação os documentos que possuem marcadores de datas recentes em relação à consulta realizada, com indexação a nível de exaustividade.

A definição da linha temporal para a recuperação dos documentos é feita de acordo com as preferências do usuário, levando em consideração as limitações de apresentação dos documentos. O objetivo é fornecer novos materiais que possam ser avaliados e reindexados, visando a atualização e renovação do conjunto de documentos disponíveis. Isso é consonante, também, com o observado por Ingwersen e Jäverlin (2005) e proposto no modelo cognitivo de Ingwersen (1996). A Figura 25 demonstra as formas de apresentação da informação recuperada, na tela apresentada ao usuário.

**Figura 25 - Interface de apresentação da informação recuperada no modelo proposto, apresentando as variedades de formas de apresentação**



Fonte: Elaborado pelo autor.

Estes quatro processos encerram as ações do golfo de execução. No golfo de avaliação, duas ações ou subprocessos podem ser observados: (1) avaliação por parte do usuário das informações recuperadas e (2) atualização dos termos vinculadores documentais.

A avaliação dos usuários permite identificar quais documentos, arquivos ou instâncias de informação são mais relevantes ou pertinentes às suas necessidades. Este processo está relacionado com os atributos contextuais do tipo linguístico, simbólico e interpretativo (LSI). Conforme apontado por Foresti, Varvakis e Godoy Viera (2016) a questão extralinguística do contexto em que um usuário se encontra se relaciona, em parte, com sua capacidade de percepção subjetiva da informação que trabalha. Essa percepção varia de indivíduo para indivíduo, o que pode influenciar na forma como ele percebe alguma informação que lhe parece ser mais relevante.

Como a captação do processo subjetivo da informação é tarefa complexa, acontecendo apenas na mente do usuário conforme apontado por Buckland e (1991) e Wilson (2006), a CI atenta-se apenas para a informação tangível, ou manipulável. Dessa forma, a avaliação do usuário imputa ao sistema essa subjetividade de forma tangível, ao escolher as informações que lhe “parecem ser” mais relevantes. Entendendo que perfis similares com contextos socioculturais e cognitivos similares tem chances de realizar escolhas similares (MINSKY, 1986), a adoção dessa ferramenta de avaliação pode atender as premissas de identificação de relevância e pertinência por parte do sistema, problemática apresentada por Ingwersen e Jäverlin (2005).

Ao avaliar um documento, arquivo ou instância de informação recuperada, o usuário atribui uma classificação de um a três, indicando o grau de relevância ou pertinência das informações (sendo 1 - pouco relevante, 2 – relevante e 3 – muito relevante ou pertinente). Esta avaliação é estendida tanto à forma quanto ao conteúdo do que é recuperado, em todos os três tipos de RI, documental, adaptada e recente. Documentos ou arquivos com uma classificação entre dois e três são considerados mais relevantes e, portanto, tem seus termos vinculadores atualizados a partir do mapa contextual daquela consulta.

Este processo de atualização ocorre no BD colunar que armazena os termos vinculadores dos documentos do repositório digital. Ao atualizar os termos vinculadores dos documentos com os dados do "mapa contextual", gerado a partir das combinações dos atributos socioculturais e cognitivos, ocorre um refinamento na representação dos documentos em termos de especificidade.

Ao adicionar elementos que conectam os documentos aos perfis específicos dos usuários, a informação é indexada não apenas pelo seu conteúdo, mas também pelo potencial de relevância para determinados usuários. Isso significa que os documentos são associados não apenas com base em seu conteúdo, mas também levando em consideração a sua adequação e valor para usuários específicos. Essa abordagem permite uma indexação mais personalizada e direcionada, facilitando a recuperação da informação relevante para cada usuário em particular.

Um documento com índices atualizados com atributos contextuais de um usuário recebe termos vinculadores de especificidade, aumentando a precisão no processo de RI. A precisão, dessa forma, neste tipo de sistema, ocorre de forma contínua, com atualização constante dos termos vinculadores dos documentos a partir dos atributos contextuais captados e reforçados pelos usuários do sistema.

À medida que ocorrem mais ciclos de busca e recuperação de informações, a quantidade de documentos potencialmente relevantes recuperados diminui em comparação com aqueles que são efetivamente pertinentes. Isso ocorre porque o processo de refino dos termos vinculadores busca torná-los mais precisos e específicos possível, levando em consideração o grau de uso da linguagem natural pelo usuário.

Quanto mais interações ocorrem, o SRI é aprimorado e ajustado com base no *feedback* do usuário. Isso resulta em uma melhor compreensão das preferências e necessidades do usuário, permitindo que os termos vinculadores sejam refinados de forma mais precisa, indo de encontro com o proposto por Vickery e Vickery (2004) e Ingwersen (1996). Dessa forma, o sistema pode se adaptar à linguagem natural utilizada pelo usuário, melhorando a correspondência entre a consulta realizada e os documentos recuperados.

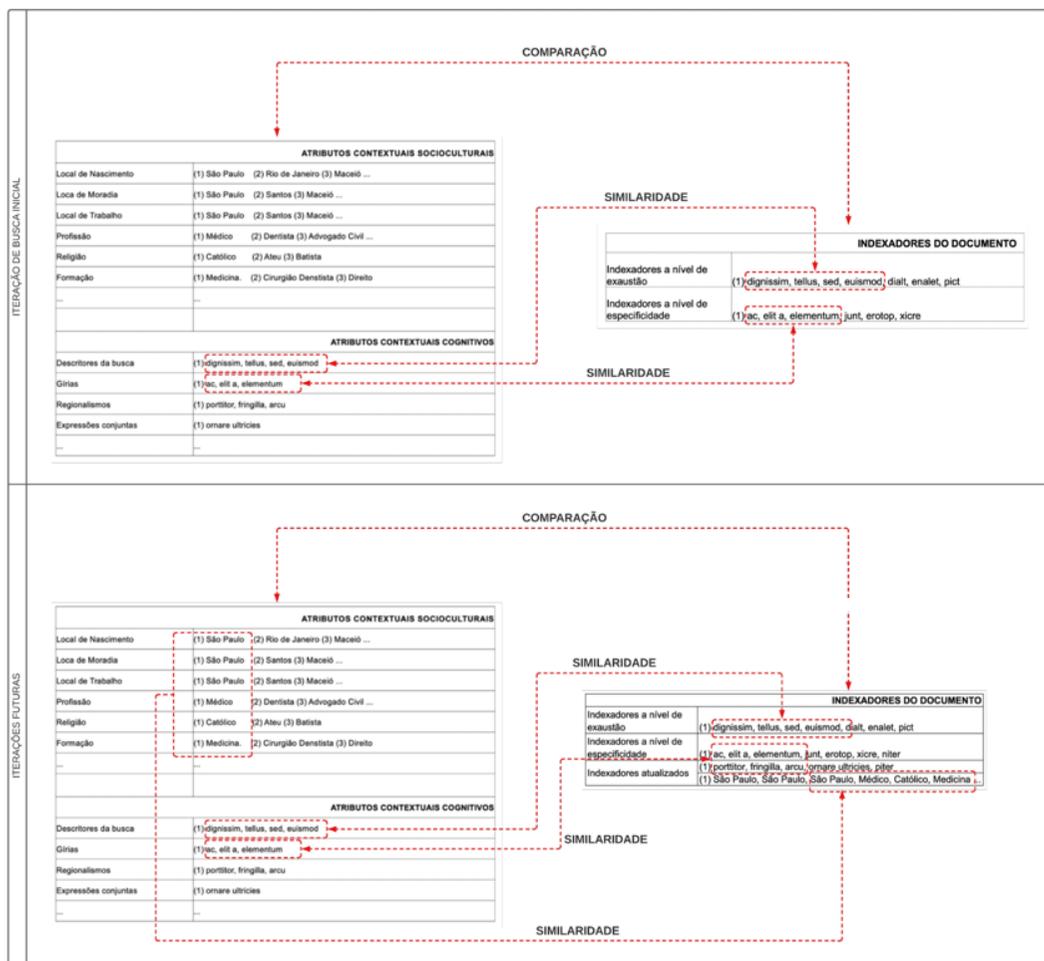
O objetivo deste processo de refinamento é maximizar a pertinência e a precisão dos documentos recuperados, fornecendo informações cada vez mais

relevantes para o usuário, à medida que o sistema aprende e se adapta às suas necessidades específicas (INGWERSEN, 1996).

Com o decorrer do tempo e o uso de ferramentas de aprendizado de máquina, os termos vinculadores que possuem pouca repetição entre os mapas contextuais e os documentos são descartados. Isso resulta em uma diminuição na exaustividade da descrição de cada documento, porém aumenta significativamente a especificidade da indexação e do referenciamento, ampliando o potencial de pertinência das informações recuperadas.

Esta abordagem aprimora a capacidade de precisão da recuperação de informações, reduzindo a necessidade de iterações subsequentes de consulta. Com a redução da revocação, ou seja, a diminuição de documentos irrelevantes sendo recuperados, aumenta-se o potencial de obter informações precisas e relevantes. A cada iteração de busca e RI, os termos vinculadores são renovados e atualizados, reforçando determinados mapas contextuais para cada perfil de usuário. A Figura 26, em tamanho maior no Apêndice 3, demonstra este processo, separadas em dois quadros: o primeiro, com comparações de mapas e termos vinculadores iniciais, e o segundo, em comparações futuras, após a atualização dos termos vinculadores dos documentos. Nota-se uma maior especificidade em sua representação, o que contribui para aumentar a precisão do processo de busca e RI, diminuindo a revocação. Isso caminha para a amplificação de pertinência dos documentos recuperados, conforme considerado como ideal por Novellino (1996).

Figura 26 – Forma de refinamento dos termos vinculadores dos documentos, a partir do processo de avaliação da informação recuperada



Fonte: Elaborado pelo autor.

É importante apontar que ao realizar nova busca, o usuário atualiza a forma como pode se expressar, o, dado o caráter de atualização dos “mapas contextuais”, as informações do sistema são atualizadas. Isso mitiga a problemática da incerteza do sistema, aumentando seu potencial para requisições dinâmicas, problemas apresentados por Ingwersen e Jäverlin (2005).

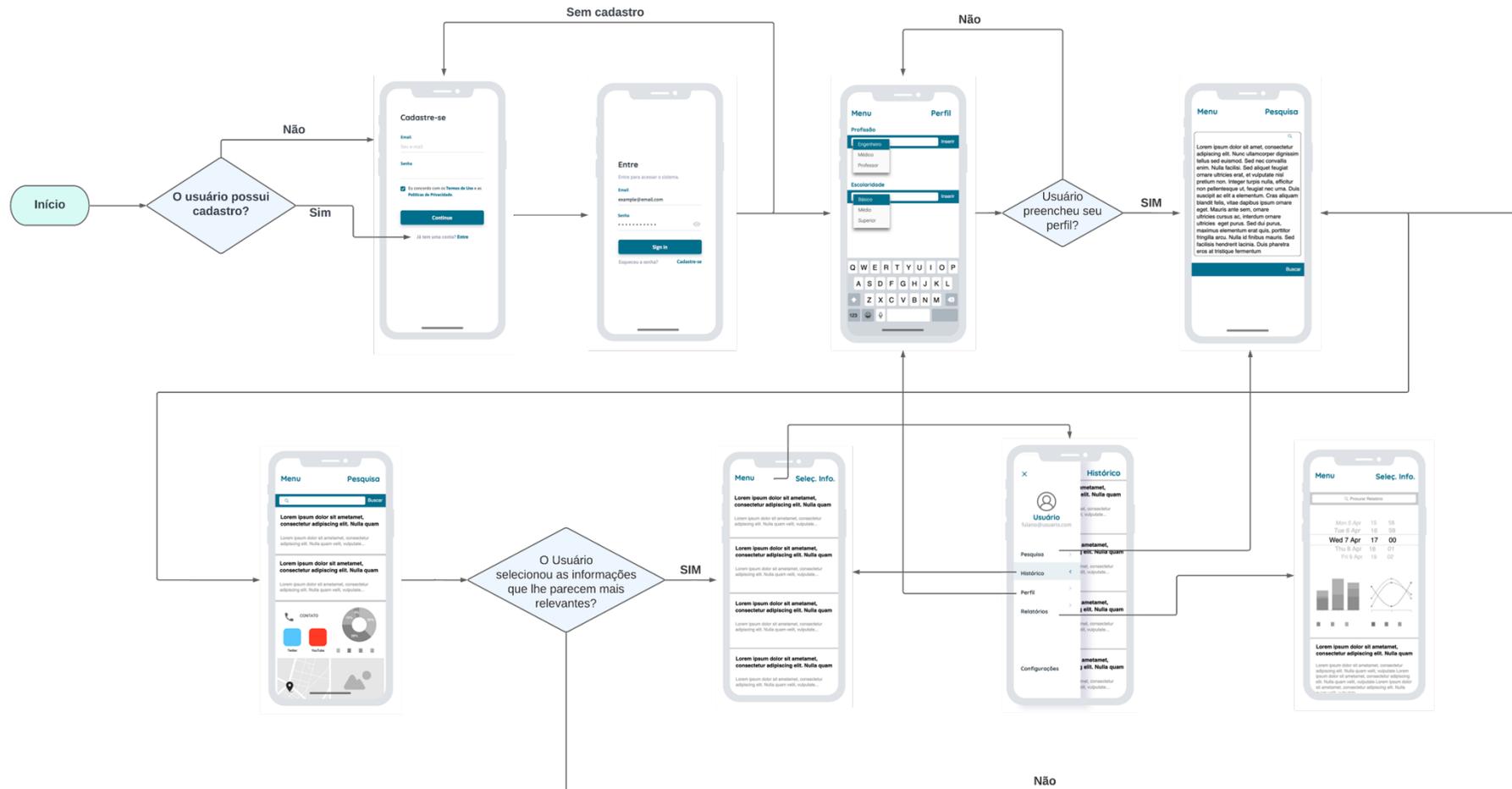
A dinamicidade do sistema depende de seu meio de mediação com o usuário, relativo à sua interface. A fim de sumarizar os processos anteriormente apresentados, segue caracterização geral da tela de acesso do usuário (Fig. 27, em tamanho maior no Apêndice 4), com explicitação de operação.

Ao acessar o aplicativo/site, o usuário é questionado sobre existência ou não de cadastro. Se houver, continua para tela de cadastro de perfil, caso tenha alguma atualização. Se não, continua para o processo de consulta.

Caso não tenha cadastro, o usuário deverá fazê-lo, para prosseguir para a tela de consulta. O seu perfil deve ser plenamente preenchido. Ao realizar sua consulta, lhe são apresentados os resultados da RI, em diversas formas e adensamentos. Ele então deve fazer a seleção dos conjuntos que lhe parecem mais relevantes. Em caso negativo de identificação de informação relevante, o processo de consulta é reiniciado. Em caso positivo, lhe é apresentada tela com histórico de busca caso queira consultar novamente iterações anteriores ou revisar a busca atual.

A partir do menu, o usuário pode acessar seu perfil para atualização, seu histórico de busca, a tela de consulta e uma área de relatórios sobre as buscas que já realizou.

Figura 27 - Diagrama de telas do fluxo de uso do SRI proposto no trabalho



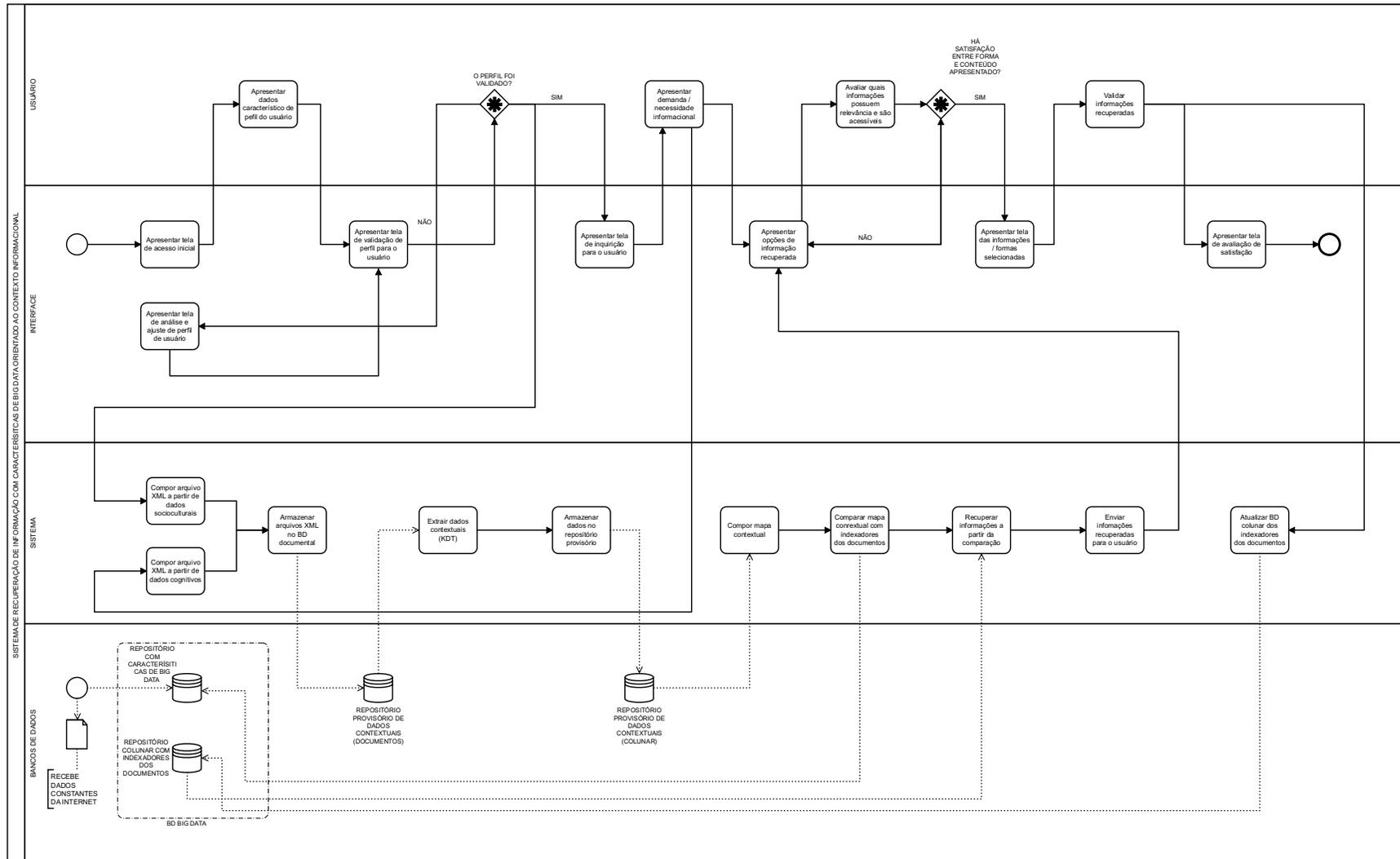
Fonte: Elaborado pelo autor.

Também, do ponto de vista do sistema, é apresentada a operação geral do modelo, considerando-se o supracitado, mostrado na Figura 28, que pode ser visto em tamanho maior no Apêndice 5.

O processo mostrado utilizando-se a técnica BMPN demonstra a divisão do modelo em quatro grandes camadas de atenção: usuário, interface, sistema e BD. Dessa forma, é possível a verificação dos processos supracitados por área de interesse.

Para demonstrar a viabilidade do modelo proposto, é apresentada uma possibilidade de aplicação deste modelo na Base de Dados da BRAPCI (Base de Referência em Informação Científica), servindo como uma prova de conceito. Nessa demonstração, o modelo é utilizado para melhorar o SRI da BRAPCI, oferecendo aos usuários uma experiência mais personalizada e eficiente na busca de conteúdos relacionados ao impacto de Jürgen Habermas no campo da CI.

Figura 28 - BPMN da operação geral do modelo proposto no trabalho



Fonte: Elaborado pelo autor.

### 4.3 Exemplo de utilização do modelo proposto

Para o modelo proposto pela presente pesquisa foi demonstrada a sua possibilidade de uso a partir de comparação com operação de “mundo real”, identificando suas potenciais vantagens e suas desvantagens em termos de uso e aplicação. Para tanto, utilizou-se como referência o SRI digital da BRAPCI. A BRAPCI é uma base de dados referencial *online*, que reúne informações bibliográficas e resumos de artigos científicos relacionados à área de CI. A BRAPCI é uma importante ferramenta para pesquisadores e profissionais interessados neste campo, pois permite o acesso a um vasto acervo de conhecimento científico.

Inicialmente projetada como um repositório focalizado na CI, a BRAPCI opera com um conjunto referencial de dados e informações estruturadas. Contudo, a amplitude de sua aplicabilidade transcende esta demarcação inicial. A BRAPCI disponibiliza referências para trabalhos que capturam o interesse de acadêmicos de diversas disciplinas, não restringindo sua utilidade aos limites da CI. Este aspecto multidisciplinar da base de dados sinaliza a necessidade de implementar mecanismos mais sofisticados que facilitem o acesso a documentos pertinentes a um espectro mais amplo de áreas acadêmicas.

Neste contexto, o algoritmo proposto na tese emerge como uma contribuição significativa, propondo uma abordagem de vinculação e acessibilidade mais diversificada. Este algoritmo vai além da indexação convencional baseada em palavras-chave, possibilitando uma exploração mais minuciosa e orientada ao contexto do uso das informações. Este ponto é corroborado pelo que apresentou Gonzalez de Gómez (2017), que enfatizou a indissociabilidade entre o usuário e seu contexto sociocultural, argumentando que as necessidades e os comportamentos informacionais devem ser compreendidos, entre outras coisas, como construções sociais situadas, entre sujeitos e meio.

O potencial de extensão da aplicabilidade da BRAPCI é particularmente notável quando considerada sua infraestrutura preexistente. As modificações propostas no trabalho não apenas preservariam a utilidade da base de dados para os pesquisadores em CI, mas também a expandiriam para abarcar uma variedade de outras disciplinas. Isto seria alcançado por meio da implementação de diferentes

níveis e métodos de acessibilidade, permitindo que a base de dados seja adaptável a variados contextos e requisitos informacionais.

Em resumo, a seleção da BRAPCI como objeto empírico desta pesquisa encontra justificativa na sua capacidade latente e no valor agregado que as inovações propostas poderiam conferir, tanto ao campo da CI quanto a outras disciplinas acadêmicas, quando da utilização da BRAPCI como base de pesquisa.

A BRAPCI é o resultado do projeto de pesquisa intitulado "Opções metodológicas em pesquisa: a contribuição da área da informação para a produção de saberes no ensino superior". Este projeto tem como objetivo fornecer suporte para estudos e propostas na área de CI, sendo implementado por meio de atividades planejadas de forma institucional. A equipe responsável pelo desenvolvimento da BRAPCI foi composta pela professora Leilah Santiago Bufrem, pelos professores Elías Sanz Casado e José Antonio Moreiro González da Universidad Carlos III de Madrid, e pela professora Wanda Maria Maia da Rocha Paranhos da Universidade Federal do Paraná.

A criação da BRAPCI teve como finalidade reunir os títulos de periódicos da área de CI e indexar seus artigos, formando, assim, uma base de dados referencial. Esta iniciativa ampliou o espaço documentário disponível para os pesquisadores, facilitando a visualização da produção na área e destacando as particularidades desse domínio científico. Atualmente, a BRAPCI conta com referências e resumos de 19.255 textos publicados em 57 periódicos nacionais impressos e eletrônicos da área de CI (BUFREM *et al.*, 2010).

Para implementar esta base de dados, foi utilizado o *ElasticSearch*, um mecanismo de busca distribuído baseado no *Apache Lucene*. O *ElasticSearch*, desenvolvido em Java e lançado sob os termos do Apache, ofereceu um mecanismo de pesquisa de texto completo e suporta vários usuários. Além disso, ele possui uma interface web HTTP e utiliza documentos JSON sem esquema para o armazenamento de dados (BUFREM *et al.*, 2010).

A ideia de criar uma base de dados para estudar a produção em CI surgiu no projeto de pós-doutorado da professora Leilah Santiago Bufrem em 1995. O objetivo era desenvolver um repertório representativo da produção científica do Brasil e da Espanha. Com a participação dos professores Elías Sanz Casado, José Antonio Moreiro González e Wanda Maria Maia da Rocha Paranhos, foi criada a base Brasil/Espanha de artigos de periódicos da área em CI (BRES) (BUFREM *et al.*, 2010).

A BRES foi planejada em 1996 e adotou o software ProCite como ferramenta de gerenciamento de referências científicas. Desenvolvido nos anos 1980, o *ProCite* permitiu a criação de bancos de dados, busca e recuperação de informações, emissão de relatórios e integração com o protocolo Z39.50. A base de dados utilizava informações provenientes de bases de dados online, CD-ROM, correio eletrônico, bibliotecas digitais e acervos de bibliotecas. Com o convênio estabelecido entre a *Universidad Carlos III* de Madrid e a Universidade Federal do Paraná em 2005, a base de dados fortaleceu-se como um projeto interinstitucional, permitindo a troca de informações e a mobilidade de professores e estudantes entre os dois países. A pesquisa realizada no âmbito desse convênio abrangeu estudos comparativos entre a produção brasileira e espanhola em Ciência da Informação (BUFREM *et al.*, 2010).

Ao longo dos anos, a base de dados BRAPCI passou por melhorias e atualizações. Em 2008, foi criado um layout e sistema de busca para a plataforma, visando facilitar o acesso e a visualização dos dados pelos usuários. Além disso, novos periódicos foram incorporados à base, ampliando o escopo e a representatividade da produção científica na área de CI.

A partir de 2012, a BRAPCI adotou o sistema *Open Journal Systems* (OJS) para gerenciar os periódicos incluídos na base. O OJS é uma plataforma de código aberto voltada para a gestão editorial e publicação de revistas científicas, oferecendo recursos para submissão, revisão, edição e publicação de artigos. Essa mudança permitiu uma maior integração entre os periódicos e a base de dados BRAPCI, agilizando o processo de inclusão dos novos artigos.

A BRAPCI é uma ferramenta essencial para a comunidade acadêmica de CI. Ela oferece acesso a referências e resumos de artigos, permitindo que pesquisadores, professores, estudantes e profissionais encontrem informações relevantes, identifiquem tendências de pesquisa e acompanhem o desenvolvimento da área. Além disso, a BRAPCI promove a visibilidade da produção científica em CI, tanto nacional quanto internacionalmente. Ao facilitar a disseminação do conhecimento e a troca de informações entre pesquisadores, ela impulsiona o progresso da área e estimula o desenvolvimento de novas pesquisas.

Como ambiente informacional digital, o site da BRAPCI possui características importantes de ambientes informacionais digitais, que possibilitam a aplicação do modelo proposto para aprimoramento. Em primeiro lugar, ela está amplamente conectada a diversos repositórios, permitindo a consulta de documentos e arquivos

de várias fontes e temas. Isso pode resultar em uma alta revocação (recuperação de informações relevantes) e baixa precisão. Em segundo lugar, devido ao seu potencial multidisciplinar, espera-se que a BRAPCI seja consultada por diferentes públicos-alvo de diferentes áreas de conhecimento, com diferentes níveis de conhecimento. Isso pode dificultar a identificação prévia dos usuários, o que deve ser considerado como um desafio.

Adicionalmente a estas duas problemáticas observadas e analisadas no processo de prova de conceito feito, outros pontos foram identificados e são apresentados a seguir. Para o desenvolvimento do processo de prova de conceito do modelo proposto, são consideradas as premissas apontadas por Balci (1998) sobre avaliação de modelos conceituais. Entende-se que modelos conceituais representam uma aplicação possível e não podem ser sujeitos aos mesmos critérios de avaliação de modelos lógicos ou físicos. Dessa forma, não se espera a análise do modelo em exaustão de simulações, mas sim, análise a partir de uma simulação conceitual que compreenda se não todos, parte significativa dos processos esperados de funcionamento do modelo.

Para a prova de conceito do modelo proposto, foram realizados testes que envolveram tanto a operação real do SRI online da BRAPCI, quanto a simulação da operação do mesmo site utilizando o modelo proposto para recuperação da informação. Devido à impossibilidade de consultar todo o acervo da BRAPCI para análise e comparação, os testes não foram realizados em um caso específico, mas sim avaliaram o potencial de funcionamento e operação dos dois ambientes.

Para tanto, é apresentado um teste de operação técnica do modelo, executando em programa suas operações principais, a fim de dar sustentação à prova de conceito. Após esta introdução, são apresentados os procedimentos de funcionamento da BRAPCI e do modelo proposto. Além disso, o modelo proposto é integrado ao sistema da BRAPCI para avaliar sua viabilidade e aplicabilidade. Ademais, são analisados os pontos de vantagem e desvantagem do uso do modelo, destacando as razões pelas quais o modelo é considerado validado.

### 4.3.1 Experimento de operações técnicas básicas do modelo

Com o objetivo de analisar o funcionamento do modelo proposto e fornecer uma base sólida para a prova de conceito, é apresentado a seguir um programa inicial que realiza as operações básicas do modelo. O programa é responsável pela criação de dados XML, captura de informações deste arquivo, criação de um BD colunar e vinculação destes dados a documentos relacionados. O programa fornecido desempenha algumas tarefas relacionadas ao modelo proposto, objetivando testar a possibilidade técnica de aplicação do modelo em ambiente de “mundo real”.

Tais tarefas estão relacionadas à criação de dados XML, captura de informações deste arquivo, criação de um BD colunar e vinculação destes dados a documentos relacionados. O programa foi desenvolvido com o objetivo de avaliar a viabilidade técnica do sistema em relação às operações mencionadas. Essas operações foram realizadas em um ambiente controlado, separado do tempo de execução real, a fim de analisar a viabilidade prática do modelo proposto em termos de aplicação.

Até o momento, o programa demonstrou uma operação eficiente, executando estes processos de maneira eficiente. No entanto, é importante ressaltar que ainda é necessário um desenvolvimento adicional, principalmente com o uso de bancos de dados documentais para o arquivamento de dados mais volumosos provenientes de arquivos XML. Além disso, a operação do sistema em um ambiente de produção em tempo real também precisa ser considerada.

No desenvolvimento do programa, foram utilizadas as linguagens de programação *Python* e KDT *stick-learn*, juntamente com o BD colunar NoSQL Cassandra. Essas tecnologias foram escolhidas para oferecer uma solução completa e eficiente para o modelo proposto. Este programa inicial representa um passo importante no processo de análise e validação do modelo, proporcionando uma base sólida para o aprimoramento e expansão futuros. O programa utiliza funções de bibliotecas como o *sklearn* para o processamento de texto, *XML.etree.ElementTree* para manipulação de XML, *XML.dom.minidom* para formatação do XML, e *cassandra.cluster* para a interação com o BD Cassandra.

Para o processamento da informação recebida em XML, o programa possui a função `salvar_em_XML(texto, ocupação, idade, formação)` que recebe um texto, ocupação, idade e formação como entrada e os armazena em um arquivo XML. Essa

função cria um arquivo XML estruturado com marcadores para cada informação fornecida.

Para extração de palavras deste arquivo XML, temos a função `extrair_palavras` (texto), que recebe um texto como entrada. Essa função utiliza técnicas de processamento de texto da biblioteca *sklearn* para extrair os substantivos, adjetivos e verbos do texto dissertativo. O resultado é uma representação vetorial dessas palavras extraídas, obtidas por meio do *CountVectorizer*, *TfidfTransformer* e *TruncatedSVD*.

Tais palavras são armazenadas em um BD colunar com a tecnologia Cassandra, criado com a função `criar_banco_dados`(substantivos, adjetivos, verbos). Nessa função, a tabela "palavras" é criada, e as palavras extraídas (substantivos, adjetivos e verbos) são inseridas no BD, juntamente com informações sobre o tipo de palavra e um link de arquivo associado.

Por fim, o programa inclui a função `vincular_arquivo_link` (palavra, arquivo\_link), que permite vincular um arquivo/link a uma palavra específica no BD Cassandra. Essa função atualiza o registro da palavra na tabela "palavras" com o arquivo/link relacionado. Assim, o programa foi capaz de criar um arquivo XML com dados específicos, extrair palavras-chave deste arquivo, armazenar essas palavras em um BD colunar (Cassandra) e vincular informações adicionais (como links de documentos) a essas palavras no BD. Isso permite uma organização estruturada e a associação de informações relevantes aos dados capturados a partir do arquivo XML.

No exemplo do programa, um texto dissertativo foi fornecido como entrada. Este texto foi salvo em um arquivo XML juntamente com informações adicionais, como ocupação, idade e formação. Em seguida, palavras-chave foram extraídas deste texto usando técnicas de processamento de texto. Estas palavras foram então armazenadas em um BD colunar usando o Cassandra. O exemplo demonstrou como o programa pode ser usado para capturar informações, organizar dados e vincular conteúdo relacionado, tornando-o útil em cenários que exigem essa funcionalidade. O código fonte pode ser visto a seguir no Quadro 6, onde são detalhados em *Python* os processos supracitados.

### Quadro 6 - Código fonte do programa de teste do modelo proposto no trabalho

```
from sklearn.feature_extraction.text import CountVectorizer
from sklearn.feature_extraction.text import TfidfTransformer
from sklearn.decomposition import TruncatedSVD
from sklearn.pipeline import Pipeline
from XML.etree.ElementTree import Element, SubElement,
tostring
import XML.dom.minidom
from cassandra.cluster import Cluster

# Função para salvar o texto do formulário em um arquivo XML
com marcadores
def salvar_em_XML(texto, ocupacao, idade, formacao):
    root = Element("texto")
    SubElement(root, "ocupacao").text = ocupacao
    SubElement(root, "idade").text = idade
    SubElement(root, "formacao").text = formacao
    SubElement(root, "conteudo").text = texto

    XML_string =
    XML.dom.minidom.parseString(tostring(root)).toprettyXML(indent=" ")
    with open("texto.XML", "w") as XML_file:
        XML_file.write(XML_string)

# Função para extrair substantivos, adjetivos e verbos do
texto dissertativo do formulário
def extrair_palavras(texto):
    pipeline = Pipeline([
        ('vect', CountVectorizer()),
        ('tfidf', TfidfTransformer()),
        ('svd', TruncatedSVD(n_components=100))
    ])
```

continua...

continuação...

#### Quadro 6 - Código fonte do programa de teste do modelo proposto no trabalho

```
palavras = pipeline.fit_transform([texto])
return palavras

# Função para criar o banco de dados Cassandra e atualizar os
índices
def criar_banco_dados(substantivos, adjetivos, verbos):
    cluster = Cluster(['localhost'])
    session = cluster.connect()

    session.execute("CREATE KEYSPACE IF NOT EXISTS meukeyspace
WITH replication = {'class': 'SimpleStrategy',
'replication_factor': '1'}")
    session.execute("USE meukeyspace")

    session.execute("CREATE TABLE IF NOT EXISTS palavras (palavra
text PRIMARY KEY, tipo text, arquivo_link text)")

    for palavra in substantivos:
        session.execute("INSERT INTO palavras (palavra, tipo,
arquivo_link) VALUES (%s, %s, %s)", (palavra, "substantivo",
""))

    for palavra in adjetivos:
        session.execute("INSERT INTO palavras (palavra, tipo,
arquivo_link) VALUES (%s, %s, %s)", (palavra, "adjetivo", ""))

    for palavra in verbos:
        session.execute("INSERT INTO palavras (palavra, tipo,
arquivo_link) VALUES (%s, %s, %s)", (palavra, "verbo", ""))

    cluster.shutdown()

# Exemplo de utilização do programa
# texto = "Lorem ipsum dolor sit amet, consectetur adipiscing
elit. Integer luctus erat non ex tristique efficitur. Vivamus
ultrices tempus vestibulum. Aliquam fringilla sapien id mi
tincidunt, nec feugiat sem bibendum. Nulla vestibulum purus ac
lectus fermentum dapibus. Donec id rutrum nulla. Nulla nec
urna eget nisi tincidunt aliquam. Quisque efficitur, est a
aliquam porttitor, lacus risus congue velit, a condimentum
lacus velit et lorem."
# ocupacao = "Engenheiro"
# idade = "30"
# formacao = "Bacharel em Engenharia"
```

continua...

conclusão...

#### Quadro 6 - Código fonte do programa de teste do modelo proposto no trabalho

```
# salvar_em_XML(texto, ocupacao, idade, formacao)
# palavras = extrair_palavras(texto)

# substantivos = ["Lorem", "ipsum", "consectetur", "luctus",
"efficitur"]
# adjetivos = ["consectetur", "ultrices", "tempus",
"fringilla", "tincidunt"]
# verbos = ["sit", "adipiscing", "luctus", "erat",
"efficitur"]

# criar_banco_dados(substantivos, adjetivos, verbos)

# Criar vínculos entre índices do Cassandra e documentos/links
def vincular_arquivo_link(palavra, arquivo_link):
# Conectar ao cluster do Cassandra
cluster = Cluster(['localhost']) # substitua 'localhost' pelo
endereço do seu cluster Cassandra
session = cluster.connect()

# Atualizar o registro no banco de dados com o arquivo/link
relacionado
query = "UPDATE palavras SET arquivo_link = %s WHERE palavra =
%s"
session.execute(query, (arquivo_link, palavra))

# Fechar a conexão com o cluster
cluster.shutdown()

# Exemplo de uso
# vincular_arquivo_link('consectetur',
'https://www.exemplo.com/arquivo1.txt')
# vincular_arquivo_link('ultrices',
'https://www.exemplo.com/arquivo2.txt')
# vincular_arquivo_link('tempus',
'https://www.exemplo.com/arquivo3.txt')
```

Fonte: Elaborado pelo autor.

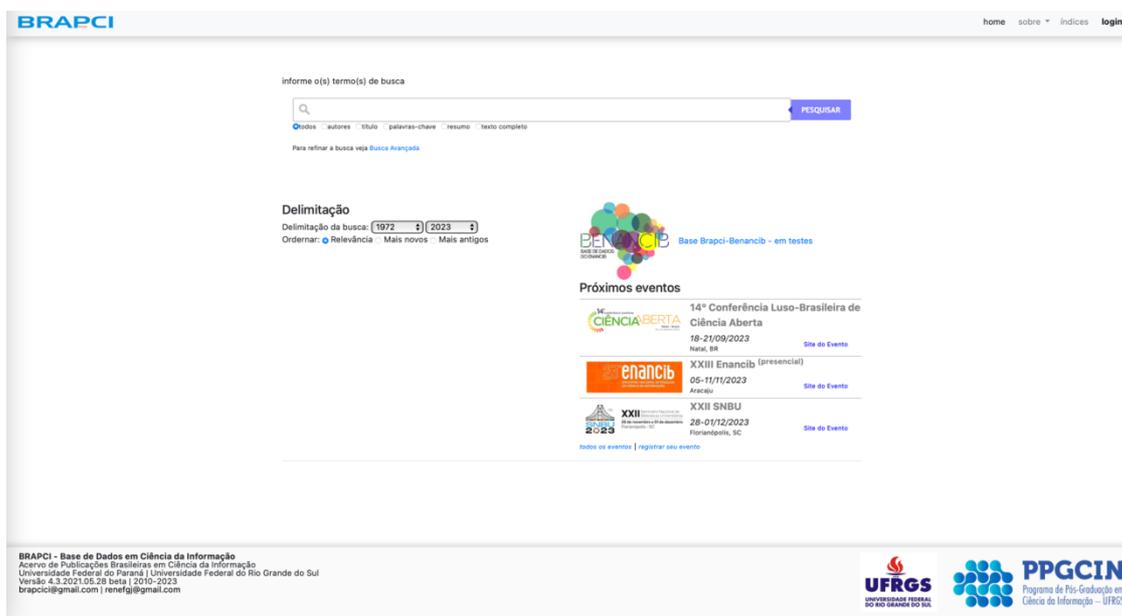
### 4.3.2 Características do site da BRAPCI

O site do SRI da BRAPCI (<https://www.brapci.inf.br>) possui menu superior com informação sobre a BRAPCI, sobre quais coleções estão indexadas no site e campo para *login*, considerado um usuário cadastrado. Para além disso, considera um campo de consulta onde podem ser solicitado documentos nos repositórios, recuperando

destes acervos a partir de autores, títulos, palavras-chave, resumos ou texto completo. Estes últimos com potencial de recuperação da informação com maior revocação e menor precisão.

É possível filtrar posteriormente a consulta os documentos e arquivos recuperados entendidos com maior relevância, bem como de forma cronológica. O site também conta com *links* os próximos eventos na área de CI, bem como links para o site da UFRGS (Universidade Federal do Rio Grande do Sul), mantenedora do SRI, desde 2010. A Figura 29 mostra o site da BRAPCI, contendo estes elementos descritos.

**Figura 29 – Printscreen da página inicial e campo de pesquisa de referências da BRAPCI**



Fonte: <https://www.brapci.inf.br> (acessado em 15/05/2023).

Ao realizar atividade de busca no site, são recuperados *links* de arquivos, artigos e documentos que estão vinculados aos descritores de busca utilizados. Utilizando os mesmos descritores ou explicitação de necessidade informação, a depender do tipo de pesquisa, obtém-se mais ou menos resultados. A RI é feita consultando-se documentos e artigos com links organizados no sistema de busca da *ElasticSearch*. Estes documentos recuperados seguem o preconizado na política de desenvolvimento do acervo da BRAPCI (GABRIEL JÚNIOR, 2011).

Para definir a política do acervo, foram considerados os seguintes critérios: a) o periódico deve ter sido avaliado em algum sistema de avaliação, como o sistema

Qualis no Brasil; b) o periódico deve estar alinhado com o escopo da base de dados; c) o periódico deve conter artigos representativos para a área em questão; d) é preferível que o periódico tenha um ciclo de vida mínimo de dois anos e/ou mais de três fascículos publicados; e) o periódico deve apresentar importância e representatividade dos autores para a área. Estes critérios buscam garantir a qualidade, relevância e consistência do acervo, bem como assegurar que a seleção de novos títulos seja embasada em critérios objetivos e alinhados com os objetivos da base de dados (GABRIEL JÚNIOR, 2011).

O cadastro feito pelo usuário, anteriormente à busca, não afeta os resultados obtidos, dessa forma, não possuindo peso no processo de RI. Assim, independentemente do perfil do usuário do sistema, as RI serão as mesmas, também, independentemente de seu contexto de uso.

A título de exemplo, considera-se o perfil de um estudante de ciências sociais, que tenha interesse em saber do impacto existente das obras de Jürgen Habermas na CI. Espera-se, dada a amplitude do repositório da BRAPCI possibilidade de recuperação da informação, neste sentido.

Duas formas podem ser consideradas para realizar a solicitação ao SRI da BRAPCI: em linguagem natural e em descritores representativos. Como observado por Ferneda (2003), é de responsabilidade do usuário a formulação de como irá compor os termos de sua consulta em um repositório informacional. No entanto, conforme Ingwersen e Jäverlin (2005) o SRI deve ser capaz de realizar a “ponte” entre estes termos e o repositório, sem prejuízo do que é recuperado ao usuário.

Considerado o exemplo, busca-se inicialmente em linguagem natural a seguinte expressão: “quais as contribuições do filósofo e sociólogo alemão Jürgen Habermas para a CI?”.

Ao realizar a busca em linguagem natural, nada foi encontrado. A Figura 30 apresenta a tela de consulta logo após a inserção da expressão.

**Figura 30 – Busca em linguagem natural no campo de referências da BRAPCI (considerando a relação CI e o filósofo Habermas)**

informe o(s) termo(s) de busca

todos  autores  título  palavras-chave  resumo  texto completo

Para refinar a busca veja [Busca Avançada](#)

**Delimitação**

Delimitação da busca:

Ordernar:  Relevância  Mais novos  Mais antigos

Nada localizado para "qual a contribuição do filósofo e sociólogo alemão Jürgen Habermas para a Ciência da Informação"

historic\_search

Data/Hora	Consulta	Tipo	Ordem	Total
2023-05-19 00:07:58	<a href="#">QUAL A CONTRIBUIÇÃO DO FILÓSOFO E SOCIÓLOGO ALEMÃO JÜRGEN HABERMA PARA A CIÊNCIA DA INFORMAÇÃO</a>	texto completo	Relevância	0

Fonte: <https://www.brapci.inf.br> (acessado em 19/05/2023).

A fim de garantir que não seja problema com a existência de documentos no acervo, o mesmo procedimento foi feito com outro filósofo, Michel Foucault. Novamente, nenhum resultado, conforme Figura 31.

**Figura 31 - Busca em linguagem natural no campo de referências da BRAPCI (considerando a relação CI e o filósofo Foucault)**

informe o(s) termo(s) de busca

todos  autores  título  palavras-chave  resumo  texto completo

Para refinar a busca veja [Busca Avançada](#)

**Delimitação**

Delimitação da busca:

Ordernar:  Relevância  Mais novos  Mais antigos

Nada localizado para "qual a contribuicao do filosofo Michel Foucault para a ciencia da informacao?\*"

historic\_search

Data/Hora	Consulta	Tipo	Ordem	Total
2023-05-19 00:11:53	<a href="#">QUAL A CONTRIBUICAO DO FILOSOFO MICHEL FOUCAULT PARA A CIENCIA DA INFORMACAO?*</a>	texto completo	Relevância	0

Fonte: <https://www.brapci.inf.br> (acessado em 19/05/2023).

Apesar de não existirem obras destes autores diretamente vinculadas à CI, existem intersecções e estudos esperados como os observados nos trabalhos de Rabello (2017) e Gonzalez de Gómez (2012), ambos vistos no referencial teórico deste trabalho, que citam estes autores. Dessa forma, entende-se que o problema está atrelado ao sistema, que não é capaz de lidar com linguagem natural.

De fato, a BRAPCI possui uma política de controle de vocabulário de busca, conforme explicado por Gabriel Júnior (GABRIEL JÚNIOR, 2011). Esta política considera diferentes formas de abordagem por meio apenas de descritores.

Os descritores utilizados na BRAPCI são classificados em diferentes categorias, de acordo com a categorização proposta por Gabriel Júnior (2011). Estas categorias são as seguintes: descritores unilexicais e polilexicais, que são empregados para definir um assunto específico; descritores cronológicos, que são utilizados para definir um período histórico; descritores geográficos, que têm o propósito de definir nomes de cidades, estados e regiões geográficas; descritores onomásticos, que são empregados para definir nomes de pessoas e instituições; outras categorias, que englobam descritores que não se enquadram nas opções anteriores; e, por fim, a categoria "não analisado", que é utilizada para avaliar descritores incorporados recentemente. A cada novo termo adicionado, este indexador é classificado em uma das categorias mencionadas, o que contribui para a organização e aprimoramento do vocabulário de busca da BRAPCI.

Assim, ao utilizar descritores, o processo de RI apresenta resultados. Utilizando-se o descritor "Habermas", tem-se o observado na Figura 32. São recuperados 36 artigos que, de fato, abordam a questão da sociologia dentro de CI. Destes, apenas 11 efetivamente lidam com a questão habermasiana, com uma precisão de 30,5%. No entanto, a recuperação é exclusivamente documental, sem juízo de valor ao público pesquisador. Dessa forma, apesar de relevante e pertinente, como alguns dos casos, a acessibilidade da informação pode ficar comprometida, dificultando a apropriação e uso por parte do usuário.

**Figura 32 – Resultado de busca de referências para o descritor “Habermas”**

informe o(s) termo(s) de busca

habermas PESQUISAR

todos  autores  título  palavras-chave  resumo  texto completo

Para refinar a busca veja [Busca Avançada](#)

**Delimitação**

Delimitação da busca: 1972 2023

Ordenar:  Relevância  Mais novos  Mais antigos

---

Selecionar Página | Selecionar Tudo Total 22

1 2

	<p><input type="checkbox"/> <b>Ética do discurso e eugenia liberal: Jürgen Habermas e o futuro da natureza humana   The ethics of discourse and liberal eugenics: Jürgen Habermas and the future of human nature</b> 2008</p> <p><i>AMARAL, Aécio</i></p> <p>Liinc em revista, n. 1, v. 4, 2008. (Redes: Articulação Entre Ciência, Tecnologia e Sociedade   Networks: Connection Between Science, Technology and Socie) 9.7540</p>
	<p><input type="checkbox"/> <b>Contribuições habermasianas para uma ética intercultural da informação: aproximações</b> 2016</p> <p><i>IULIANELLI, Jorge Atilio Silva</i></p> <p>Logeion: filosofia da informação, n. 1, v. 3, p. 29-41, 2016. {} 9.6532</p>
	<p><input type="checkbox"/> <b>Justificação e a ação de informação no contexto da pragmática virtual   Justification and action information in the context of pragmatics virtual</b> 2010</p> <p><i>GRACIOSO, Luciana de Souza</i></p> <p>Liinc em revista, n. 2, v. 6, 2010. (Dossiê Linguagem, Informação e Nova Dinâmica Social Contemporâneas I) 9.1474</p>
	<p><input type="checkbox"/> <b>Brevíssimas considerações sobre a busca de uma Ação Comunicativa entre o pensamento de Habermas e a Diplomática Arquivística Contemporânea</b> 2007</p> <p><i>RONDINELLI, Rosely Curi</i></p> <p>Arquivística.net, n. 2, v. 3, 2007. (Ensaio) 8.6974</p>
	<p><input type="checkbox"/> <b>Cartadas do jogo informacional : a perspectiva dual da informação como matriz do mundo sistêmico e do mundo vivido</b> 2013</p> <p><i>MEDEIROS, José Washington de Morais; FIDELIS, Marli Batista</i></p> <p>Informação &amp; Sociedade: Estudos, n. 1, v. 23, 2013. (Relato de Pesquisa) 8.6974</p>

Fonte: <https://www.brapci.inf.br> (acessado em 19/05/2023).

Mesmo com a possibilidade de filtro das informações, essa é feita apenas a nível cronológico. A nível de relevância, são recuperados os documentos com maior número de termos vinculadores atendidos pela busca. A fim de mitigar essa problemática, o SRI deveria, conforme explicitado por Ingwersen e Jäverlin (2005), Vickery e Vickery (2004) e Ingwersen (1996) ser adaptativo ao usuário, apresentando para ele informações acessíveis, adaptadas à sua capacidade de interpretação. Ao ter recuperação apenas documental, problema apontado por Ingwersen e Jäverlin (2005), a interpretação do recuperado limita-se à cognição do usuário.

A fim de mitigar esta problemática, Ingwersen (1996) propõe processo de retroalimentação do sistema, onde o usuário, em condição de insatisfação, dá retorno para o sistema a fim de obter nova recuperação mais alinhada com sua cognição e necessidade informacional. Este não é o caso no site da BRAPCI, tendo em vista que não há espaço para *feedback* direto para retroalimentação do sistema. Há, no entanto, dada a política de indexação do site, sistema de interação assíncrona. O sistema

permite ao usuário marcar e selecionar trabalhos representativos com base em sua leitura. Essa seleção forma um novo conjunto de documentos, chamado de corpus, que pode ser armazenado, impresso ou enviado por e-mail. Além disso, o sistema oferece funcionalidades para gerar indicadores de análises, auxiliando na avaliação e exploração dos resultados obtidos (GABRIEL JÚNIOR, 2011).

Dado que não há uma ligação direta entre o usuário e os gestores da base de dados, o sistema disponibiliza três formulários para o usuário interagir: o primeiro permite relatar sua experiência com o sistema, o segundo possibilita informar qualquer inconsistência na base de dados ou quebra de links, e o terceiro permite entrar em contato com o grupo E3PI responsável pela base. Todas as interações e envios são analisados pela equipe de gestão, que toma as medidas necessárias para corrigir eventuais problemas e fornecer um retorno adequado ao usuário (GABRIEL JÚNIOR, 2011).

Pode-se utilizar novos descritores, porém as RI serão potencialmente diferentes. Há a opção de busca avançada, que apresenta métodos de utilização da área de consulta do site para refinar o processo de busca. Em busca composta, conforme proposto pela área de busca avançada, vista na Figura 33, podem ser utilizados os operadores *booleanos* AND e OR.

## Figura 33 – Métodos de busca avançada incorporadas atualmente ao sistema da BRAPCI

- **Busca Composta**

Na busca composta o sistema insere automaticamente o elemento booleano OR entre os termos, recuperando apenas os registros que tenham a ocorrência.

ex: Bibliometria Citação

Forma de busca do sistema: Bibliometria OR Citação

---

Para delimitar a busca, onde ocorra a ocorrência dos termos é necessário a inclusão do elemento booleano AND entre os termos, recuperando apenas os registros que tenham as ocorrências indicadas.

ex: Bibliometria Citação

Forma de busca do sistema: Bibliometria AND Citação

---

Em buscas com mais de um termo, pode-se atribuir pesos diferentes para cada um deles, definindo maior "prioridade para um deles", para isso, atribua a inficação "^" e o peso atribuído

ex: Biblioteca AND Universit\*^10

O sistema recupera todos os termos biblioteca e universit\*, porém atribui um peso maior para a "universit\*", colocando com maior relevância nos resultados.

Fonte: <https://www.brapci.inf.br> (acessado em 19/05/2023).

Utilizando-se o operador AND para verificar cruzamentos dos tipos exclusivos, tem-se (Fig. 34), a partir dos descritores “informação” e “Habermas” o seguinte:

**Figura 34 – Busca de referências dos descritores “Habermas” e “Ciência da Informação” utilizando o operador avançado “AND”**

informe o(s) termo(s) de busca

Para refinar a busca veja [Busca Avançada](#)

**Delimitação**

Delimitação da busca:

Ordernar:  Relevância  Mais novos  Mais antigos

---

Selecionar Página | Selecionar Tudo Total 11

	<input type="checkbox"/> <b>Ética do discurso e eugenia liberal: Jürgen Habermas e o futuro da natureza humana   The ethics of discourse and liberal eugenics: Jürgen Habermas and the future of human nature</b> <i>AMARAL, Aécio</i>	2008
	Liinc em revista, n. 1, v. 4, 2008. (Redes: Articulação Entre Ciência, Tecnologia e Sociedade   Networks: Connectiom Between Science, Technology and Socie)	11.9496
	<input type="checkbox"/> <b>Justificação e a ação de informação no contexto da pragmática virtual   Justification and action information in the context of pragmatics virtual</b> <i>GRACIOSO, Luciana de Souza</i>	2010
	Liinc em revista, n. 2, v. 6, 2010. (Dossiê Linguagem, Informação e Nova Dinâmica Social Contemporâneas  )	11.8663
	<input type="checkbox"/> <b>Mediação da informação e ação comunicativa Habermasiana</b> <i>FIDELIS, Marli Batista; GOMES, Henriette Ferreira</i>	2022
	Logeion: filosofia da informação, v. 9, p. 91-111, 2022. (Artigo)	9.5375
	<input type="checkbox"/> <b>A presença de Habermas na ciência da informação no Brasil</b> <i>FREIRE, Isa Maria</i>	2022
	Logeion: filosofia da informação, v. 9, p. 222-237, 2022. (Artigo)	9.5235
	<input type="checkbox"/> <b>A Produção Acadêmica no Brasil sobre Ciência da Informação: um estudo a partir da Teoria do Conhecimento de Habermas</b> <i>VITAL, Luciane Paula; SILVEIRA, Murilo Artur Araújo da; PINHO, Fabio Assis; CORREIA, Anna Elizabeth Galvão Coutinho; FELL, André Felipe Albuquerque</i>	2014
	Em Questão, n. 1, v. 20, p. 127-150, 2014. (Artigo)	9.2494
	<input type="checkbox"/> <b>Administração da informação para a inovação em uma organização; Gestión de la información para la innovación en una organización</b> <i>CARVALHO, Lidiane dos Santos; LIMA, Clovis Ricardo Montenegro de</i>	2013
	Informação@Profissões, n. 2, v. 2, p. 1-20, 2013. (Comunicação)	9.1918

Fonte: <https://www.brapci.inf.br> (acessado em 19/05/2023).

Neste ciclo são recuperados os 11 documentos indicados anteriormente. A utilização destes operadores *booleanos* aumenta a precisão de busca, refinando o processo de RI a partir da indexação. No entanto, novamente, não há possibilidade de retorno para o sistema de forma direta e a recuperação é exclusivamente documental.

Para fins de comparação, na próxima seção, são apresentadas as operações da BRAPCI considerando o modelo proposto.

### 4.3.3 Integração do modelo de SRI proposto ao site da BRAPCI

Intenciona-se adicionar o modelo proposto ao site da BRAPCI, dotando-o de mais uma opção de método para RI, a partir de seu banco de repositórios. Não se intenciona substituir completamente seu método original de projeto. Para tanto, são necessárias adaptações no sistema atual do site da BRAPCI, no que tange sua interface, sistema e BD. Seis são as ações necessárias que devem ser feitas para que haja integração do modelo proposto com o sistema da BRAPCI, e são descritas a seguir:

Adaptação da estrutura de dados: como visto, o site da BRAPCI é uma base de dados referencial, ou seja, aponta para os documentos e arquivos em *links* de acesso de outras bases de dados, contendo as informações potencialmente relevantes para um usuário, a partir dos termos de busca utilizados. Assim o faz utilizando da tecnologia *ElasticSearch*. Seu funcionamento é dividido em cinco principais componentes: índices, *shards*, nós, cluster e consulta. Os quatro primeiros processos estão relacionados ao gerenciamento e armazenamento dos links para as bases de dados referenciadas. Os dados são organizados em índices, que são coleções lógicas de links para documentos relacionados em outras bases de dados. Estes índices possuem descritores iniciais que indicam, de forma indexada, o conteúdo possível dos links.

Cada índice possui um nome único e é dividido em *shards*. Os *shards* representam subconjuntos dos links do índice e são unidades de armazenamento e processamento independentes baseados no Apache Lucene. Estes *shards* podem ser distribuídos em diferentes nós do cluster, permitindo uma distribuição equilibrada dos dados e um processamento paralelo mais eficiente.

Cada nó é uma instância do *ElasticSearch* em um servidor e é responsável por armazenar e processar um ou mais *shards* do índice. Os nós se comunicam entre si para compartilhar dados, coordenar a indexação e a busca, e garantir a alta disponibilidade dos links para as bases de dados. Juntos, estes nós formam um cluster, que é um conjunto de nós que trabalham em conjunto para gerenciar e processar os links. Eles se comunicam por meio de um protocolo interno, compartilhando informações sobre os *shards* e coordenando as operações de busca e indexação. Isso possibilita uma escalabilidade horizontal e um aumento na capacidade de processamento. A escalabilidade horizontal envolve adicionar mais

servidores ao ambiente de computação para que eles possam trabalhar em conjunto e compartilhar a carga de trabalho. Isso permite que o sistema ou aplicação lide com um aumento significativo na demanda, mantendo o desempenho e a disponibilidade.

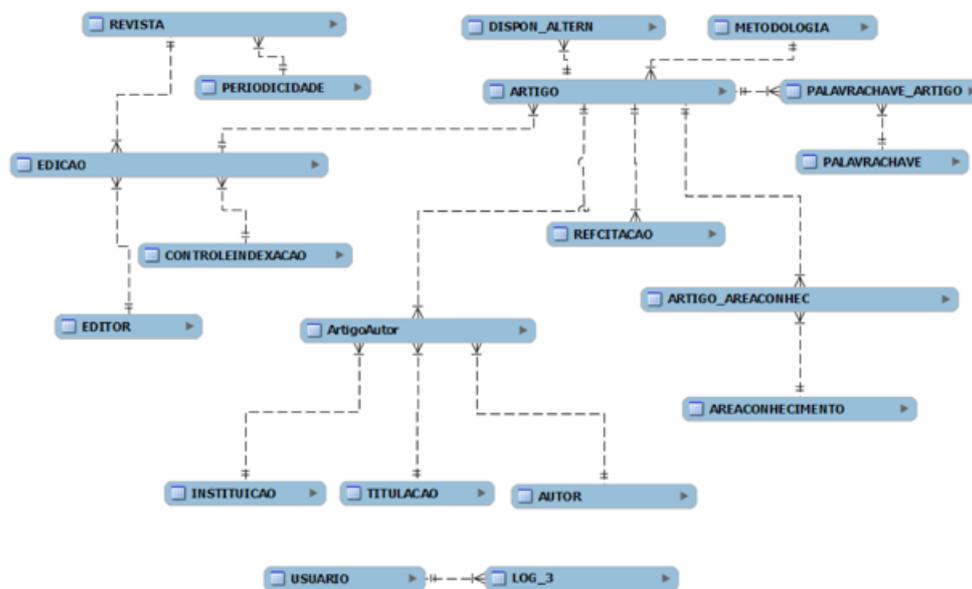
A consulta é o processo pelo qual os links para as bases de dados são pesquisados no *ElasticSearch*. As consultas são expressas em formato JSON<sup>21</sup> e podem incluir critérios de pesquisa, filtros, ordenação e outros parâmetros. O *ElasticSearch* distribui a consulta para os *shards* relevantes no cluster, buscando os links correspondentes nas bases de dados referenciadas. Os resultados são retornados de forma rápida e eficiente, permitindo que os usuários obtenham os links para as bases de dados desejadas com base em suas consultas. A busca é a última etapa do processo, em que os usuários podem recuperar eficazmente os links e acessar as informações nas bases de dados referenciadas.

As bases de dados referenciais são organizadas em estrutura SQL. Como visto, uma base de dados SQL organiza as informações em tabelas para recuperação posterior através de método de consulta. A consulta do *ElasticSearch*, distribuída nos *shards*, identifica em quais tabelas os *links* para os documentos relacionados estão salvos, para recuperação à um usuário. O modelo de relacionamento de dados pensados para a BRAPCI pode ser visto na Figura 35, retirada de Gabriel Júnior (GABRIEL JÚNIOR, 2011).

---

<sup>21</sup> JSON (*JavaScript Object Notation*) é um formato de dados leve e de fácil compreensão utilizado para estruturar e transmitir informações entre sistemas. Ele é baseado na sintaxe do *JavaScript*, mas pode ser utilizado com várias linguagens de programação (HARRISON, 2019).

**Figura 35 - Modelo de relacionamento de dados (BRAPCI)**



Fonte: Gabriel Júnior (2011, p. 86).

A relação da base de dados da BRAPCI, em configuração SQL, permite a busca nas tabelas que referenciam o documento a nível de autor, nome, revista e nome do artigo. Essas tabelas são indexadas ao processo de busca do *ElasticSearch* através de índices. Cada índice é relativo à uma tabela do SQL e é vinculado à uma linha que contém as informações dos *links* dos repositórios.

A utilização de índices dentro do *ElasticSearch* para organização dos *links* das tabelas do SQL é semelhante à etapa inicial proposta no modelo, que envolve a identificação de termos vinculadores iniciais para construção das recuperações de informação preliminares, antes da atualização com termos contextuais. Dessa forma, ao integrar o modelo proposto na base de dados referencial da BRAPCI, é considerada a utilização dessa estrutura já existente de indexação para criação destes termos vinculadores iniciais entre *links* de documentos termos de busca. No entanto, o modelo proposto considera a vinculação destes termos vinculadores e documentos através de BD colunar, onde cada conjunto de termos vinculadores é vinculado à um índice dentro do BD. Dessa forma, é prevista também a integração de um BD colunar para armazenamento destes índices.

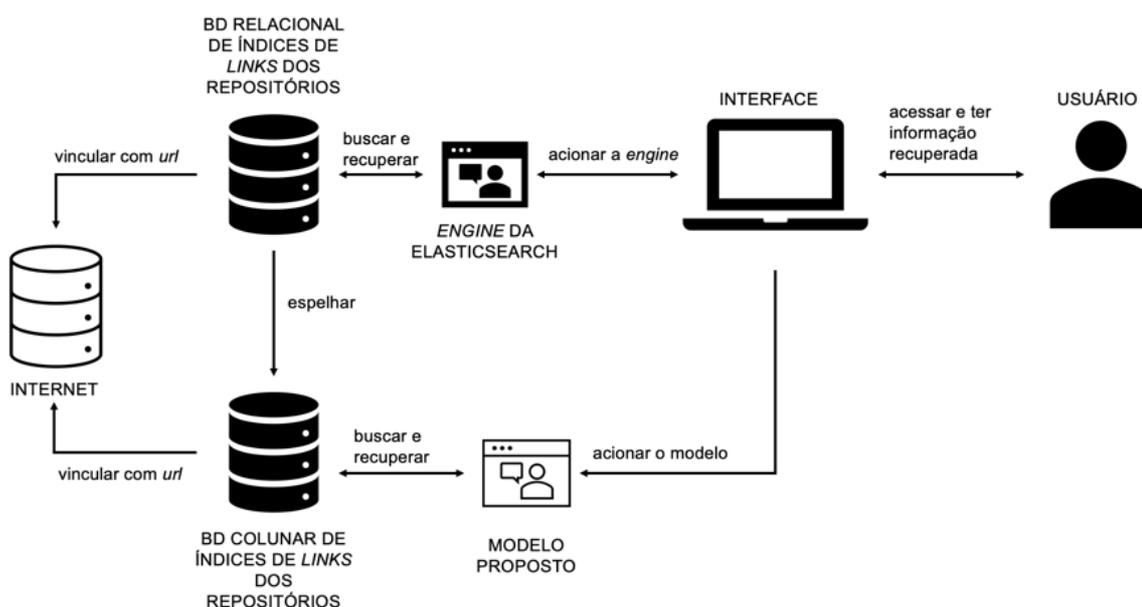
O BD relacional que contém os termos e os valores das linhas da tabela, que identificam os documentos, precisa ter um espelho em um BD colunar. Neste BD colunar, as informações são ajustadas de modo que cada *link* possua um índice,

considerando cada conjunto de termos como termos vinculadores. Isso permite que o BD colunar seja atualizado com atributos contextuais, de acordo com o que é proposto pelo modelo.

Ademais, textos como resumos e outras estruturas mais complexas podem ser armazenados integralmente, levando em consideração que a comparação de termos vinculadores no modelo proposto utiliza ferramentas como o GPT, capazes de identificar elementos específicos dentro dos textos. A base de dados da BRAPCI, em BD relacional do tipo SQL, deve ser espelhada em um BD colunar do tipo NoSQL. Deste BD colunar então é utilizado como base para classificação inicial de termos vinculadores para os links, adicionando-os em índices iniciais.

A Figura 36 demonstra essa adaptação de vinculação, demonstrando como a base de dados da BRAPCI precisará ser vinculada para a utilização do modelo proposto. Será realizado um espelhamento do BD relacional em um BD colunar, para que o modelo proposto tenha acesso aos índices inicialmente pensados no projeto da BRAPCI. A partir daí, o modelo pode realizar as vinculações iniciais entre índices e links de documentos, prosseguindo com seu procedimento de trabalho previsto. O processo, conforme demonstrado, não afeta o uso com a *engine* da *ElasticSearch*.

**Figura 36 – Proposta de integração do modelo apresentado com o sistema da BRAPCI**



Fonte: Elaborado pelo autor.

No modelo, o documento ou arquivo carrega apenas o seu índice, que está vinculado ao BD colunar. Isso facilita a atualização dos atributos termos vinculadores dos documentos, favorecendo a utilização de atributos contextuais futuros para aprimorar sua recuperação. Assim, quando é feita uma consulta, são analisados e comparados os valores de vinculação salvos neste BD colunar. Ao serem encontradas similaridades, são recuperados documentos com o mesmo índice dos valores com similaridade, de forma mais célere. No caso do site da BRAPCI, o mesmo pode ser considerado, porém com os *links* de acesso dos documentos, salvos em índices do BD colunar.

Apenas o BD colunar é atualizado com os dados de indexação dos *links* dos documentos, a partir dos atributos contextuais. O BD relacional, inicialmente projetado para o site original da BRAPCI, é mantido para consultas ao acervo usando os métodos originais do site.

Sistema de cadastro de usuários: inicialmente, o preenchimento de cadastro de usuário não afeta o processo de recuperação da informação no site da BRAPCI. Como a identificação de atributos socioculturais do usuário é essencial para o processo de RI a partir de atributos contextuais de uso da informação, este processo deve ser adaptado para a captação de atributos socioculturais do usuário. Para tanto, o sistema de cadastro deve conter, para além das informações básicas do usuário, formulário de preenchimento de atributos socioculturais, conforme apresentado na seção 4.2.

A interface do formulário pode ser criada utilizando *HTML*, que é uma linguagem de marcação de hipertexto. Essa linguagem pode ser integrada a várias linguagens de programação, como PHP, *Python*, JavaScript, entre outras. Estas linguagens são amplamente utilizadas para integração com servidores e bancos de dados, especialmente no desenvolvimento web. As linguagens permitem a conexão com bancos de dados por meio de extensões específicas, que possibilitam a execução de consultas, inserções, atualizações e exclusões de dados tanto em bancos de dados SQL quanto NoSQL. Também, essas linguagens têm a capacidade de coletar dados e organizá-los em documentos do tipo XML, necessários para posterior análise KDT para composição dos mapas contextuais.

Estes arquivos XML devem ser armazenados em um BD documental, que também deve ser considerado em previsão de adaptação do sistema da BRAPCI.

Também, ademais do BD documental, é necessária previsão de BD colunar provisório, para além do BD colunar de índices. É nele que os atributos contextuais socioculturais e cognitivos serão salvos para o desenvolvimento do mapa contextual.

Adaptação do sistema de consulta: o sistema de consulta da BRAPCI é composto por tela em página do *site*, com campo de pesquisa, desenvolvida em HTML. Para integração com o *ElasticSearch*, é utilizada tecnologia JavaScript. Ao inserir um conjunto de descritores de busca, o algoritmo em JavaScript relaciona estes termos com os salvos nos índices do *ElasticSearch*, recuperando os documentos para o usuário.

O modelo proposto captura atributos contextuais cognitivos durante o processo de descrição e expressão do problema do usuário. Para tanto, ele precisa organizar essas informações em documento XML para posterior análise e armazenamento de termos retirados em BD colunar, com índices para cada iteração de busca e RI. Os dois bancos de dados (documentos e colunar) descritos anteriormente fazem essa função. Cabe, no campo de pesquisa, utilizando de linguagem PHP, *Python*, JavaScript ou outra, desenvolver ferramenta para captura da descrição e organização em arquivo XML, para posterior registro em BD de documentos.

Também, similarmente ao processo anterior, os documentos XML são analisados via KDT para identificação de termos interessantes para composição dos atributos que serão colocados no mapa contextual da iteração atual de busca e RI. Os dados coletados e armazenados nestes BDs podem ser distribuídos nos *shards* do *Elasticsearch* para agilizar o processo de RI. Cada *shard* é um índice independente que contém uma parte dos dados do índice completo. O *ElasticSearch* distribui automaticamente os *shards* pelo cluster, aproveitando a capacidade de processamento e armazenamento de vários nós. Assim, é possível uma distribuição inteligente dos links e dados, agilizando o processo de consulta.

Sistema de funcionamento do site: para o pleno funcionamento do sistema apoiado ao site da BRAPCI, é necessário implementar diferentes componentes. A captura de atributos socioculturais e cognitivos requer o desenvolvimento de um algoritmo, o qual pode ser realizado com linguagens de programação como *Python*, PHP, Ruby ou Node.js, integrado à interface web do site da BRAPCI. Os atributos capturados devem ser armazenados em bancos de dados documentais do tipo NoSQL, como *MongoDB* ou *ElasticSearch*, utilizando as bibliotecas ou extensões adequadas da linguagem escolhida.

A análise em KDT, visando a composição de mapas contextuais, pode ser realizada utilizando algoritmos específicos, como *scikit-learn*, implementados em linguagens como *Python*, R ou Java, que funcionam bem quando da utilização de bancos de dados NoSQL. A recuperação de informações e comparação com documentos existentes envolve o uso dos mapas contextuais resultantes da análise em KDT. As linguagens mencionadas anteriormente são adequadas para implementar essa lógica.

A atualização dos termos vinculadores dos documentos é realizada com base nos resultados da comparação e no *feedback* dos usuários. A lógica para essa atualização pode ser implementada na linguagem escolhida, interagindo com os bancos de dados relacionais ou documentais utilizados.

A apresentação das informações recuperadas, bem como suas diferentes formas de apresentação, deve considerar a utilização de tecnologias GPT em paralelo com as tecnologias de HTML, PHP e JavaScript, que podem fazer uso do poder generativo do GPT para apresentação textual e imagética em níveis simplificados, intermediários e complexos.

É fundamental considerar a escalabilidade, a eficiência e a manutenibilidade do sistema ao selecionar a linguagem e as tecnologias apropriadas. No caso da presente pesquisa, que envolve um modelo conceitual, essa preocupação não se destaca, porém, é necessário destacar a importância dessas considerações para o sucesso do projeto, em caso de implementação efetiva.

Essas implementações não consideram a desativação do sistema atual, mas sim, sua complementação em paralelo. As tecnologias apresentadas podem ser consideradas como complementares às já utilizadas no site, demandando de infraestrutura de serviço específica como mais espaço em servidores e BD para operação dos programas do sistema.

Sistema de RI com múltiplas formas: para apresentar os diferentes níveis de explicação da informação recuperada pelo GPT dentro da estrutura atual do site da BRAPCI, é possível utilizar tecnologias como HTML, PHP e JavaScript. Com o HTML, é possível organizar a informação recuperada em diferentes níveis de explicação, utilizando elementos como parágrafos, listas ou tabelas.

O PHP, por sua vez, é uma linguagem de programação que possibilita o processamento de informações no lado do servidor. Com o *PHP*, é possível integrar o GPT ao sistema e realizar as tarefas de geração de texto nos diferentes níveis de

explicação. Por exemplo, ao receber uma solicitação do usuário, o *PHP* pode acionar o GPT e obter as explicações simplificadas, intermediárias ou complexas, que serão exibidas na página web de informações recuperadas da BRAPCI.

Além disso, o JavaScript é uma linguagem de programação que roda no navegador do cliente, permitindo a criação de elementos dinâmicos e interativos na página web. Com o JavaScript, é possível adicionar funcionalidades como botões ou menus, que permitem ao usuário organizar o nível de explicação desejado. Dessa forma, o conteúdo exibido na página pode ser alterado de acordo com a escolha do usuário, proporcionando uma experiência mais personalizada para além da recuperação puramente documental da BRAPCI.

Ao combinar as tecnologias mencionadas anteriormente, é possível criar uma interface web interativa que apresenta a informação recuperada pelo GPT em diferentes níveis de explicação. Os usuários terão a capacidade de explorar as explicações simplificadas, intermediárias e complexas, de acordo com suas necessidades e preferências. Para garantir o armazenamento adequado das instâncias dos documentos apresentados de formas diferentes no site da BRAPCI, é recomendado utilizar um BD NoSQL como estrutura documental.

O site da BRAPCI, por possuir uma estrutura referencial, não trabalha diretamente com os documentos. Portanto, é necessário armazenar as instâncias dos documentos em um BD NoSQL, que servirá como uma base de dados adicional para consultas. Essas instâncias devem considerar os termos vinculadores atualizados pelos mapas contextuais das iterações de busca e RI. Dessa forma, o BD NoSQL permitirá consultas eficientes e precisas, levando em consideração as informações geradas pelo GPT e as atualizações dos termos vinculadores.

Essa abordagem garante que as diferentes formas de apresentação da informação, com seus respectivos níveis de explicação, estejam disponíveis para consulta no site da BRAPCI. Os usuários poderão explorar essas informações, utilizando a interface web interativa, e o BD NoSQL armazenará as instâncias dos documentos para suportar as consultas futuras, levando em consideração os termos vinculadores atualizados pela busca e RI realizadas pelos usuários.

Sistema de *feedback* de usuário: após o processo de recuperação de informação, é necessário aplicar à página as informações recuperadas uma estrutura de avaliação em três níveis, que vão do 1 ao 3, classificando as informações de acordo com sua relevância e pertinência. Essa estrutura de avaliação permitirá determinar

quais documentos e instâncias de informação recuperada devem ter seus termos vinculadores atualizados, visando aprimorar a recuperação da informação futura dos documentos dentro da BRAPCI.

Para desenvolver a interface e o sistema proposto, várias ferramentas podem ser utilizadas. Algumas delas são as tecnologias web, como HTML, CSS e JavaScript, que possibilitam a criação de interfaces interativas e amigáveis para o usuário. O HTML é usado para estruturar a página e exibir as informações recuperadas, enquanto o CSS permite estilizar a página com cores, fontes e layouts atrativos. O JavaScript é empregado para adicionar interatividade, como a seleção dos níveis de relevância e pertinência.

Além disso, é necessário desenvolver um sistema de atualização dos índices do BD colunar que contém os termos vinculadores dos documentos e suas instâncias. Isso envolve o uso de tecnologias de desenvolvimento de sistemas, como PHP, para processamento no lado do servidor, e bancos de dados SQL ou NoSQL, para armazenar e gerenciar os dados relevantes.

Também é importante mencionar a necessidade de desenvolver uma página de avaliação do sistema, que não existe atualmente na estrutura da BRAPCI. Essa página permitirá aos usuários avaliar a qualidade da Recuperação da Informação e, em caso de insatisfação, reiniciar o processo de recuperação da informação. Novamente, as tecnologias apresentadas, como HTML, CSS, JavaScript e PHP, podem ser utilizadas para implementar essa funcionalidade.

#### **4.3.2 Operação do modelo de SRI proposto na BRAPCI**

Ao acessar o campo de pesquisa baseada em contexto no modelo proposto integrado ao repositório da BRAPCI, o usuário pode realizar consultas sobre o impacto de Jürgen Habermas no campo da CI, conforme exemplo anterior. Para utilizar o SRI proposto, o usuário é solicitado a preencher seu perfil pessoal por meio de um formulário que requer informações contextuais, como nome, local de residência, ocupação, entre outros.

Este processo de preenchimento do perfil diferencia-se do site "real" da BRAPCI, uma vez que considera o perfil individual do usuário como um indicativo da forma e acessibilidade das informações recuperadas. É importante ressaltar que o

cadastro do usuário na BRAPCI atualmente não afeta o processo de recuperação da informação.

Considera-se a utilização do site da BRAPCI como base para inserção do modelo, como campo de pesquisa contextual. Este processo, proposto no modelo apresentado, é complementar aos outros processos de busca vistos no site da BRAPCI. Dessa forma, as telas apresentadas a seguir utilizam como base o site da BRAPCI, e os pontos de alteração são pontuados e detalhados.

A Figura 37 mostra a tela de acesso do sistema proposta para utilização no modelo, considerando a interface de cadastro do usuário, com formulário. O acesso é feito na mesma localização atual de *login*, porém de forma atualizada e considerando os potenciais atributos socioculturais. Atualmente, não há tela para cadastro do usuário, apenas campo para *login*, onde é feito acesso para o registro do usuário, em caso de interesse.

**Figura 37 – Proposta de interface de cadastro de usuário na BRAPCI, utilizando o modelo proposto no trabalho**

ID	Perfis	Perfis	Local	Status	Ações
1	job_title	Data	job opening		Edit
2	applicant_name	Data	applicant name		Edit

Fonte: Elaborado pelo autor.

Ao fornecer as informações do perfil, o usuário contribui para o preenchimento de um documento XML com seus dados, que será utilizado para o levantamento dos seus atributos socioculturais por meio de técnicas de KDT. Este documento XML é

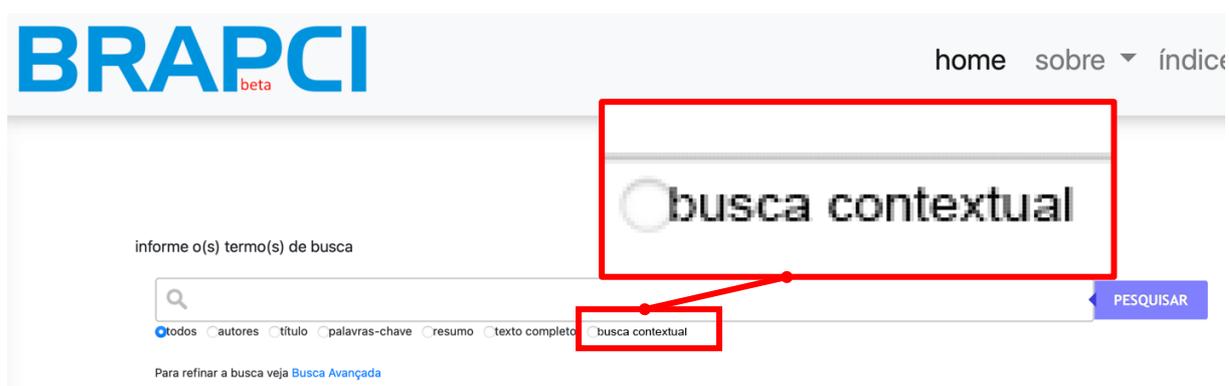
armazenado em um BD NoSQL do tipo documental, enquanto os dados extraídos deste documento XML são salvos em um BD colunar, juntamente com um índice provisório relacionado à iteração atual de busca e RI.

Estas operações são essenciais para a análise e composição dos mapas contextuais no sistema da BRAPCI. A interface de cadastro, como visto, é desenvolvida em linguagem de marcação *HTML* com integração ao BD através de linguagens PHP, JavaScript ou outras. O processo de KDT para coleta dos atributos contextuais socioculturais é feita através de contagem estatística utilizando-se *Python* ou R.

Estes atributos estão relacionados com o perfil do usuário. No caso do exemplo, como estudante de ciências sociais, o usuário pode destacar em qual série se encontra, em qual universidade de qual região. Pode indicar também se exerce alguma ocupação ou ainda estágio entre outras características descritivas.

Após preencher o formulário que coleta os atributos socioculturais, o usuário é redirecionado para a página de consulta, onde pode expressar seu problema de pesquisa sobre o impacto de Jürgen Habermas no campo da CI, utilizando tanto descritores de busca quanto linguagem natural. Ela se mantém similar à atual. É acrescentado, no entanto, campo para busca contextual, para utilização junto do modelo proposto. Este campo não existe como campo de seleção no site, atualmente. Ele é acrescentado para operação do site junto do código do modelo. Este campo novo acrescentado pode ser visto na Figura 38, que usuário deve selecionar.

**Figura 38 – Campo de seleção para utilização do processo de busca contextual, conforme modelo apresentado no trabalho**



Fonte: Elaborado pelo autor.

Diferentemente do site convencional da BRAPCI, o modelo proposto oferece flexibilidade de pesquisa, permitindo ao usuário, estudante de ciências sociais, indicar não apenas a expressão do problema, mas também seu nível de conhecimento prévio sobre o assunto e a forma como acredita ter mais chances de acessar e compreender as informações recuperadas.

Utilizando a técnica KDT, o sistema é capaz de captar tanto os termos de busca presentes no texto apresentado pelo usuário quanto os atributos contextuais cognitivos expressos. Através de treinamento prévio do KDT, expressões relacionadas à busca, como nomes, relações e temas, assim como características específicas, como nível de conhecimento e idioma, podem ser capturadas e consideradas para a composição do mapa contextual, juntamente com os atributos socioculturais.

Por exemplo, os atributos contextuais da pesquisa do estudante podem conter expressões como “pouco conhecimento prévio” ou “conhecimento avançado do tema”. Além disso, podem incluir expressões como “quem foi Jürgen Habermas” ou “entendendo que Jürgen Habermas foi um sociólogo”. Essas expressões, de acordo com as regras de treinamento do KDT, são capturadas e utilizadas como atributos para potencial reindexação dos links da BRAPCI, visando estabelecer conexões entre o perfil do usuário, sua necessidade informacional, seu estado anômalo de conhecimento e os links dos documentos ou suas instâncias. Isso, obviamente, após avaliação e validação das informações recuperadas por parte do usuário.

Assim como no processo de captura dos atributos socioculturais, o texto de consulta fornecido pelo usuário é armazenado em um arquivo XML, que é salvo no BD de documentos integrado ao sistema da BRAPCI. Posteriormente, este texto é utilizado para análise e atualização do BD colunar provisório com os dados contextuais cognitivos recuperados, utilizando a técnica KDT. Essas informações, juntamente com o índice da iteração atual de busca e RI, podem ser aproveitadas no processo de atualização dos termos vinculadores contextuais dos links dos documentos da BRAPCI, aprimorando a precisão e relevância dos resultados obtidos.

Internamente, o sistema armazena no BD provisório as relações dos atributos contextuais socioculturais e cognitivos do estudante de ciências humanas, que incluem informações sobre a temática relacionada a Habermas na CI. Isso resulta na criação de um mapa contextual da interação de busca e recuperação de informação atual.

Utilizando programas desenvolvidos em *Python*, Java ou outra linguagem compatível com a conexão a bancos de dados colunares, o sistema compara os atributos dos links salvos na BRAPCI com os atributos levantados no mapa contextual da interação, recuperando informações que possuam vínculos nos bancos de dados colunares.

Também utilizando a tecnologia GPT, o sistema propõe informações em diversos níveis para a apreciação do usuário. Essa recuperação é apresentada em uma tela de seleção ao usuário, utilizando linguagens de marcação HTML integradas ao PHP ou JavaScript para avaliação.

A tecnologia generativa refere-se a sistemas ou algoritmos que têm a capacidade de criar ou gerar conteúdo original, como texto, imagens, áudio e até mesmo vídeos. Estes sistemas são projetados para aprender padrões e estruturas a partir de conjuntos de dados de treinamento e, em seguida, usar este conhecimento para criar novas instâncias de conteúdo de forma autônoma (ZANT; SCHOMAKER; KOUW, 2012).

A tecnologia generativa é frequentemente baseada em modelos de aprendizado de máquina, como redes neurais artificiais. Estes modelos são treinados com grandes volumes de dados, permitindo que eles aprendam a capturar características e distribuições do conteúdo existente. Com base neste aprendizado, os sistemas generativos podem produzir novas amostras que se assemelham ao conteúdo original, mas com certa variabilidade ou originalidade (ZANT; SCHOMAKER; KOUW, 2012).

Dessa forma, após a recuperação de informação, o usuário tem acesso a uma variedade de recursos, como artigos, documentos e diferentes formas de apresentação da informação sobre o impacto de Jürgen Habermas no campo da CI. Isso permite que o usuário selecione as fontes e formatos que considere mais relevantes e pertinentes para suas necessidades. A Figura 39 demonstra a tela de informações recuperadas, onde apresenta documentos com vinculação direta e formas diferentes produzidas pela tecnologia GPT. Atualmente, no site da BRAPCI, apenas o campo esquerdo com os documentos recuperados é vista. Com a integração do modelo proposto, é acrescido o campo direito de múltiplas visualizações da informação.

**Figura 39 – Proposta de interface de apresentação de informações recuperadas, conforme modelo apresentado no trabalho**

The image displays a proposed interface for presenting recovered information. It is divided into two main sections: a list of search results on the left and a detailed view of a selected article on the right.

**Search Results List (Left):**

- Literacia da informação em contexto universitário (recensão)**  
PINTO, Patrícia Rosado  
Cadernos BAD (Portugual), n. 1, p. 132-135, 2017. (Leitura) 7.0845
- Métodos estratégicos gerais para estudos da informação no contexto virtual**  
CRUZ, Ruleandson do Carmo  
Informação em Pauta, n. Especial, v. 4, p. 10-25, 2019. (Artigo) 7.0711
- A importância do contexto na Ciência da Informação**  
FORESTI, Fabricio; VARVAKIS, Gragório; VIERA, Angel Freddy Godoy  
Biblios (Peru), n. 72, p. 1-21, 2018. (Original) 7.0351
- Contexto informacional e o fenômeno da hospitalidade**  
PRESSER, Nadi Helena; SILVA, Eli Lopes da; WERLANG, Elisabete  
Informação & Sociedade: Estudos, n. 3, v. 26, 2016. (Relato de Pesquisa) 7.0249
- Socialização da informação: aportes da teoria da ação comunicativa**  
SILVA, Junia Guimarães e; MARINHO, Inaldo Barbosa  
Ciência da Informação, n. 3, v. 1996. (Comunicação) 6.9672
- Os sistemas de classificação bibliográfica como interface biblioteca/usuário**  
GIGANTE, Maristela Cid  
Ciência da Informação, n. 2, v. 25, 1996. (Artigo) 6.9193
- Accountability: novo conceito para a Ciência da Informação?**

**Article Detail View (Right):**

**TEXTO I**  
A partir da leitura dos textos motivadores seguintes e com base nos conhecimentos construídos ao longo de sua formação, redija texto dissertativo-argumentativo em modalidade escrita formal da língua portuguesa sobre o tema "A persistência da violência contra a mulher na sociedade brasileira", apresentando proposta de intervenção que respeite os direitos humanos. Selecione, organize e relacione, de forma coerente e coesa, argumentos e fatos para defesa de seu ponto de vista.

**TEXTO II**  
Nos 30 anos decorridos entre 1980 e 2010 foram assassinadas no país acima de 92 mil mulheres, 43,7 mil só na última década. O número de mortes nesse período passou de 1.353 para 4.465, que representa um aumento de 230%, mais que triplicando o quantitativo de mulheres vítimas de assassinato no país.

**TEXTO III**

**TEXTO IV**  
Com base na Lei Maria da Penha, mais de 330 mil processos foram instaurados apenas nos juizados e varas especializadas

**332.216** processos que envolvem a Lei Maria da Penha chegaram, entre setembro de 2006 e março de 2011, aos 52 juizados e varas especializadas em Violência Doméstica e Familiar contra a Mulher

**237 mil**

**TIPO DE VIOLÊNCIA RELATADA**

Violência Física	Violência psicológica	Violência moral	Violência sexual	Violência patrimonial	Câncer privado	Trafico de pessoas
51,00%	31,00%	9,00%	4,00%	1,00%	2,00%	2,00%

Fonte: Elaborado pelo autor.

O sistema proposto no modelo, assim, vai além da simples indexação por palavras-chave e recupera diferentes formas de conteúdo e densidade de informações, como é atualmente no site referencial da BRAPCI.

O modelo utiliza a relação entre os termos vinculadores e os termos de busca para recuperar documentos específicos. Essa abordagem oferece ao usuário diferentes maneiras de visualizar a mesma informação, indo além dos documentos já recuperados pela indexação. O usuário pode avaliar essas diferentes formas e selecionar aquelas que melhor atendam ao seu perfil e interesse. Essa flexibilidade de apresentação difere do site convencional da BRAPCI e é uma vantagem para usuários que buscam informações com diferentes níveis de detalhamento em relação aos artigos tradicionais.

Em alguns casos, o usuário pode precisar apenas de um conjunto específico de informações contidas em um artigo. A capacidade de reestruturação do sistema permite atender a essa demanda, indo além da recuperação exclusivamente documental realizada pela BRAPCI. Para fornecer diferentes níveis de explicação da informação recuperada pelo modelo GPT, é possível utilizar tecnologias como HTML, PHP e JavaScript. Por meio do HTML, é possível organizar a informação recuperada em diferentes níveis de detalhamento, utilizando elementos como parágrafos, listas ou tabelas.

Portanto, se um estudante de graduação busca informações introdutórias sobre o papel de Habermas na CI, o modelo proposto pode gerar um resumo com

infográficos explicativos, fornecendo uma compreensão dos conceitos básicos. Além disso, além dos documentos recuperados, o modelo pode sugerir formas e possíveis relações entre os documentos para apresentar a um pesquisador mais avançado, permitindo análises aprofundadas. Essas diferentes formas de apresentação são baseadas nos conteúdos dos documentos já recuperados e apresentados, representando instâncias dos processos de recuperação de informação.

Com as informações recuperadas, o usuário tem a possibilidade de avaliar aquelas que considera adequadas por meio da interface do sistema. O usuário tem a opção de qualificar as informações recuperadas em três níveis: nada relevante, relevante e pertinente. Essa qualificação visa melhorar o SRI ao treinar as vinculações entre os mapas contextuais e os termos vinculadores de documentos. Após a apresentação das informações, o sistema solicita ao usuário a seleção das formas e densidades de informação que parecem ser mais úteis e acessíveis. Esta seleção representa um processo de retorno e *feedback*, que não está presente no site real da BRAPCI. A Figura 40 mostra em detalhe o campo considerado na BRAPCI, em conjunto com o modelo proposto, para avaliação da informação recuperada.

**Figura 40 – Proposta de campo de avaliação da informação recuperada, conforme modelo apresentado no trabalho**

Selecionar Página | Selecionar Tudo

1 2 3 4 5 6 7 8 9 10 >

Total 2981

A partir da leitura dos textos motivadores seguintes e com base nos conhecimentos construídos ao longo de sua formação, redija texto dissertativo-argumentativo em modalidade escrita formal da língua portuguesa sobre o tema **"A persistência da violência contra a mulher na sociedade brasileira"**, apresentando proposta de intervenção que respeite os direitos humanos. Selecione, organize e relacione, de forma coerente e coesa, argumentos e fatos para defesa de seu ponto de vista.

**TEXTO I**

Nos 30 anos decorridos entre 1980 e 2010 foram assassinadas no país acima de 92 mil mulheres, 43,7 mil só na última década. O número de mortes neste período passou de 1.363 para 4.485, que representa um aumento de 230%, mais que triplicando o quantitativo de mulheres vítimas de assassinato no país.

INQUÉRITO 2012 - Mapa de Violência 2013 (Análise) - Instituto Brasileiro de Geografia e Estatística (IBGE) | Disponível em: www.ibge.gov.br | Acesso em: 24 Jun. 2013.

**TEXTO II**

**TIPO DE VIOLÊNCIA RELATADA**

Tipo de Violência	Porcentagem
Violência física	51,63%
Violência psicológica	31,05%
Violência moral	9,49%
Violência sexual	2,00%
Violência patrimonial	0,94%
Outros	1,78%
Outros	0,24%

IBGE. Secretaria de Políticas para as Mulheres. Estatísticas 2014. Caderno de Estatísticas e Mulher. Disponível em: www.ibge.gov.br | Acesso em: 24 Jun. 2013 (000101010).

**TEXTO III**

**TEXTO IV**

**O IMPACTO EM NÚMEROS**

Com base na Lei Maria da Penha, mais de 330 mil processos foram instaurados apenas nos juizados e varas especializadas

**332.216** processos que envolveram a Lei Maria da Penha chegaram, entre setembro de 2006 e março de 2011, aos 52 juizados e varas especializadas em Violência Doméstica e Familiar contra a Mulher existentes no País. O que resultou em:

- 33,4%** de processos julgados
- 9.715** prisões em flagrante
- 1.577** prisões preventivas decretadas

**58** mulheres e **2.777** homens enquadrados na Lei Maria da Penha estavam presos no País em dezembro de 2010. Ceará, Rio de Janeiro e Rio Grande do Sul não constam desse levantamento feito pelo Departamento Penitenciário Nacional.

**237 mil** relatos de violência foram feitos ao Ligue 180, serviço telefônico da Secretaria de Políticas para as Mulheres

**Sete** de cada **dez** vítimas que telefonaram para o Ligue 180 afirmaram ter sido agredidas pelos companheiros

Fonte: Conselho Nacional de Justiça, Departamento Penitenciário Nacional e Secretaria de Políticas para as Mulheres. Disponível em: www.observatorio.org.br | Acesso em: 24 Jun. 2013 (000101010).

Fonte: Elaborado pelo autor.

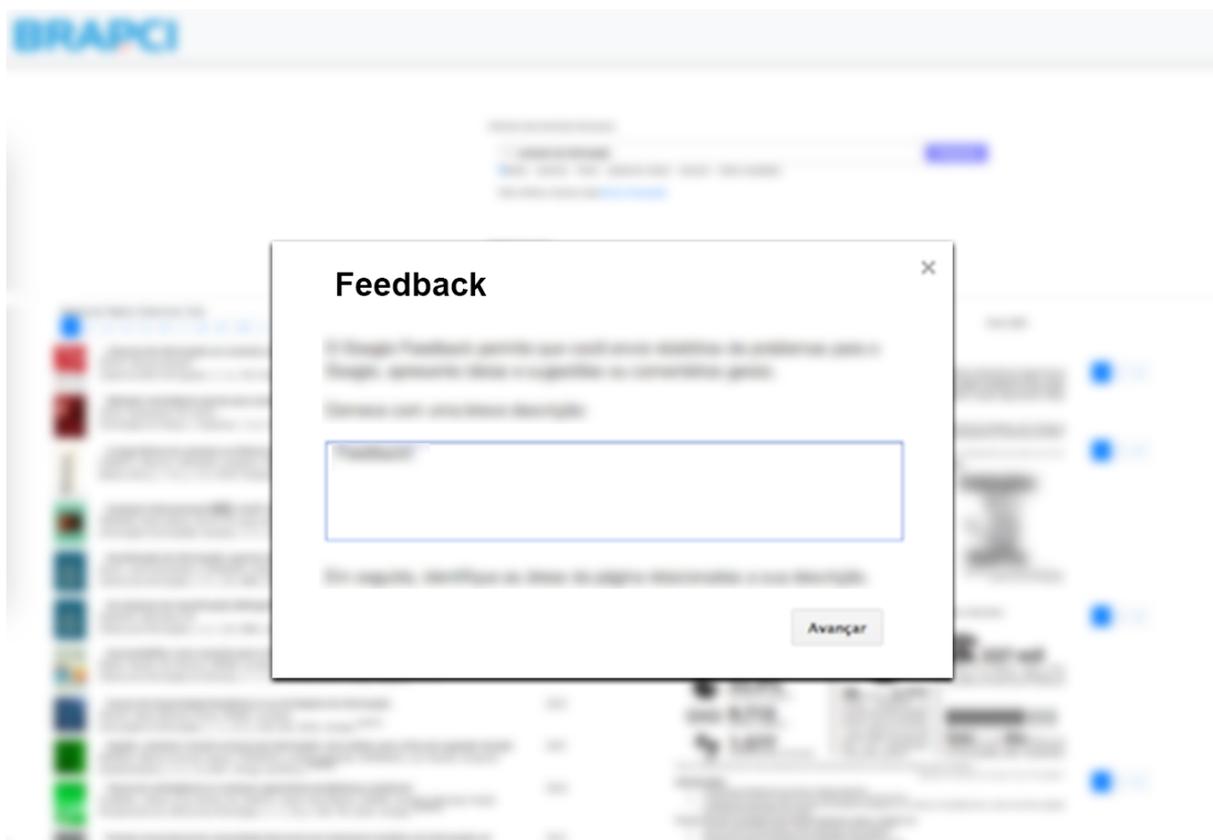
Esta avaliação dá para o sistema informações para indicar as atualizações dos termos vinculadores dos documentos ou instância de informação na BRAPCI. Essa atualização considera, no BD colunar de termos vinculadores, atualizar os índices com os atributos retirados dos processos de captação contextual sociocultural e cognitiva.

Ao atualizar um *link* recuperado, com novos termos vinculadores, ele caminha para uma maior precisão, relacionando de forma mais direta o usuário do problema dentro da BRAPCI. Dessa forma, o estudante usado no exemplo terá seu perfil, bem como sua necessidade informacional, a nível de indexação, atribuídos à um índice de link documental ou de instância de documento, refinando a especificidade de forma posterior de determinado documento.

Se o usuário não encontrar nenhuma informação pertinente, o sistema irá apresentar novamente a tela de consulta, permitindo que o usuário descreva seu

problema de forma diferente para buscar novas informações. Após avaliar as informações apresentadas, o usuário tem a opção de avaliar sua experiência, que será registrada como um log para uso futuro. A Figura 41 apresenta a interface de *feedback* do processo.

**Figura 41 – Proposta de interface de *feedback*, conforme modelo apresentado no trabalho**



Fonte: Elaborado pelo autor.

Este processo, de forma cíclica, reforça os novos termos vinculadores dos documentos do repositório, fazendo com que a informação recuperada seja mais refinada. Para tanto, dados de todos os usuários do sistema (socioculturais e cognitivos) são usados para refino dos termos vinculadores de documentos, formas e adensamentos de conteúdos.

Ao atualizar o documento com dados de atributos contextuais, de forma inclusos, o sistema vincula perfil, cognição e acesso, juntamente com conteúdo da informação. Isso, como visto, amplifica o potencial de recuperações pertinentes e com

precisão. O site da BRAPCI, com processo avançado, pode recuperar informações pertinentes, no entanto, em uma única forma.

Do ponto de vista de adaptação de sistema, isso converge com o proposto por Ingwersen (1996). O sistema é atualizado conforme *feedback* do usuário, se adaptando à sua realidade cognitiva e física. Ao longo do tempo, com atualização dos termos de indexação dos documentos a partir da seleção de informações recuperadas pelos usuários, o sistema refina o processo de RI, caminhando para um maior nível de satisfação do usuário, a nível de pertinência da informação.

Além disso, conforme proposta do modelo, pode-se recuperar informações mais recentes por meio da busca em dados recém-armazenados, o que proporciona um conjunto de informações atualizadas e relevantes para o processo de busca e recuperação da informação, evitando o problema do ajuste excessivo. Ao vincular apenas palavras-chave, o site da BRAPCI, atualmente, carece de instrumentos para trabalhar com informações atualizadas, tendo seu repositório atualizado de forma menos célere.

#### **4.3.4 Considerações sobre a integração do modelo ao site da BRAPCI**

Inicialmente, seis pontos são observados de forma patente quando da comparação dos dois processos de RI:

- a) Capacidade de trabalho em linguagem natural: O sistema proposto é capaz de compreender e processar a linguagem natural, captando problemas e nuances presentes nas consultas dos usuários. Isso melhora a busca por informações relevantes, adicionando elementos de indexação que consideram aspectos cognitivos, como localização, idioma e jargões. Em comparação, um sistema como a BRAPCI tem limitações na compreensão da linguagem natural, o que pode resultar em respostas menos precisas e completas.
- b) Realização de novas buscas de forma diferente: O sistema proposto permite a realização de buscas de maneiras diversas, o que proporciona resultados variados. Isso é útil para explorar diferentes perspectivas e encontrar informações adicionais relevantes. Em contraste, um sistema informacional digital, similar à BRAPCI, pode ter uma abordagem de busca mais limitada e menos flexível, resultando em resultados menos diversificados.

- c) Recuperação de informações em diferentes formas: O sistema proposto é capaz de recuperar informações em diferentes formatos, adaptando-se à capacidade de acessibilidade do usuário. Isso garante que a informação seja apresentada de forma adequada e compreensível, levando em consideração as necessidades específicas de cada usuário. Por outro lado, um sistema com informacional digital, similar à BRAPCI, pode fornecer informações em um formato padronizado, que pode não atender a todos os usuários adequadamente, como no caso de artigos científicos.
- d) Capacidade de *feedback* do usuário: O sistema proposto permite que os usuários forneçam *feedback*, contribuindo para a melhoria contínua do sistema. Essa capacidade de adaptação e personalização com base no *feedback* do usuário é uma vantagem significativa em relação a um sistema como a BRAPCI, que pode ter dificuldade em incorporar os *insights* dos usuários em tempo real.
- e) Apresentação de novas informações de alta revocação: O sistema proposto apresenta informações atualizadas com alta revocação para avaliação pelos usuários. Isso permite que eles explorem uma variedade de opções e acessem informações relevantes que podem ter passado despercebidas. Um sistema como a BRAPCI pode não oferecer este nível de abrangência e variedade de informações atualizadas.

De forma geral, ao considerar os contextos de uso da informação, o sistema se mostra mais adaptativo e flexível ao usuário, conforme considerado por Ingwersen e Jäverlin (2005), Vickery e Vickery (2004) e Ingwersen (1996). Isso representa a maior vantagem e contribuição do modelo proposto, tendo em vista que a rigidez de SRIs de ambientes informacionais digitais, a nível computacional, impossibilita a identificação de seu usuário.

Esta adaptação aprimora o processo de recuperação da informação pois aproxima tanto em forma quanto em conteúdo o usuário, suas necessidades e o estoque informacional, adaptando-o ao seu potencial usuário, a partir de refino dos termos vinculadores de seus documentos. De fato, este é um dos maiores problemas observados atualmente em repositórios informacionais digitais, que possuem grandes volumes de potenciais informações, porém, com pouca capacidade de transformá-las em ativos de valor para usuários específicos (BARLOW, 2013).

O modelo proposto visa aprimorar os processos de recuperação de informações em tais ambientes, estabelecendo uma conexão mais eficiente entre a forma e o conteúdo de uma consulta e o perfil do usuário. Isso é alcançado por meio do aprimoramento dos termos vinculadores documentais e das instâncias de informação.

O modelo tem como reconhecido que a recuperação de informações não se resume apenas a palavras-chave inseridas em uma busca, mas também leva em consideração os atributos contextuais do usuário e do problema informacional em questão. Compreender o contexto sociocultural, profissional e pessoal do usuário permite que o sistema personalize os resultados, oferecendo informações relevantes e compreensíveis para suas necessidades específicas.

Além disso, o modelo proposto também considera aspectos relacionados à forma, conteúdo e acessibilidade da informação. Reconhece-se que diferentes usuários possuem diferentes preferências e capacidades de acessibilidade, e, portanto, o sistema é capaz de adaptar a apresentação da informação de acordo com essas necessidades. Isso pode incluir diferentes formatos de apresentação, como texto, áudio ou visual, bem como diferentes níveis de densidade e complexidade, para atender às preferências e capacidades dos usuários.

O modelo proposto também leva em consideração o simbolismo presente na informação. Isto implica que o sistema não se limita apenas ao conteúdo textual, mas também considera os símbolos, jargões e nuances presentes na informação. Estes elementos desempenham um papel importante na captura das sutilezas e peculiaridades de um determinado campo de conhecimento ou área temática, possibilitando uma busca mais precisa e relevante.

Em resumo, o modelo proposto busca aprimorar os processos de recuperação de informações digitais, considerando a importância dos atributos contextuais, a personalização dos resultados, a adaptabilidade na forma e acessibilidade da informação, bem como o simbolismo envolvido. Isto resulta em uma experiência de busca mais eficiente e satisfatória para os usuários, proporcionando informações relevantes e compreensíveis para suas necessidades específicas. Nota-se, assim, que o modelo opera no sentido de aprimorar os processos de RI em ambientes com informacionais digitais no sentido de melhor vincular a forma e o conteúdo de uma consulta com seu usuário, a partir do aprimoramento de termos vinculadores documentais e de instâncias da informação. Desta forma, o modelo pode ser

considerado conceitualmente funcional, ao atender à essas especificações indicadas no problema de pesquisa da tese.

## **CAPÍTULO 5 – DISCUSSÃO**

Neste capítulo, "Discussão", é retomado o problema de pesquisa e hipótese formulada, estabelecendo uma conexão direta entre ambos e os resultados alcançados por meio do desenvolvimento do modelo proposto. Além disso, é analisada a viabilidade e a eficácia do modelo, considerando os aportes teóricos utilizados no desenvolvimento da pesquisa.

Ao realizar essa avaliação, consideram-se as implicações teóricas e práticas do modelo proposto. Foram identificados aspectos favoráveis do modelo, que demonstraram ser eficazes na solução do problema de pesquisa e no teste da hipótese formulada. Da mesma forma, são discutidas as possíveis limitações ou desafios encontrados durante sua implementação e utilização.

Por meio dessa análise, procurou-se fornecer uma visão abrangente sobre a contribuição do modelo proposto em relação ao problema de pesquisa e à hipótese do trabalho para a CI. Ao considerar tanto os aspectos positivos quanto as eventuais restrições, buscou-se obter uma compreensão completa do potencial de aplicação do modelo e das áreas em que pode ser aprimorado ou refinado.

### **5.1 Abordagem do modelo em relação ao problema de pesquisa e à hipótese**

Retomando o problema de pesquisa: *como promover a recuperação de informações em ambientes digitais, tendo em vista que estes ambientes são diferentes dos "tradicionais" sistemas de informação?*

Segundo Barlow (2013), ambientes informacionais digitais geralmente apresentam desafios na recuperação de informações de qualidade para usuários específicos, devido à sua arquitetura construtiva. Este ponto é reforçado por De Mauro *et al.* (2016), que indicaram que os processos de recuperação de informações nestes ambientes dependem fortemente de métodos matemáticos e estatísticos. Estes métodos visam organizar a grande quantidade de dados e informações gerados nestes ambientes - que normalmente são altamente dinâmicos e variáveis -, para fornecer ao usuário um conjunto de informações que agreguem valor à tomada de decisões (DAVENPORT, 2014). Este processo normalmente não leva em conta a efetiva necessidade informacional de um usuário (EKBIA *et al.*, 2015), restringindo-se à organização e apresentação de informações oriundas de processos de mineração de dados, conhecidos como *analytics*.

A fim de mitigar este problema, o presente trabalho teve como hipótese que a utilização do contexto de uso da informação poderia ajudar a melhorar o processo de recuperação de informações em tais ambientes. Entende-se o contexto de uso da informação como um conjunto de informações que está relacionado às características dos usuários (como, onde e quando estão, seus aspectos sociais e cognitivos, entre outros) e aos atributos das condições de uso da informação (como forma, linguagem, conteúdo e simbolismo, entre outros). Para analisar esta possibilidade, foi proposto um modelo conceitual, aplicado em um estudo de caso.

Neste sentido, foi importante levar em consideração os fluxos informacionais presentes nestes ambientes, abrangendo os processos de seleção, entrada, classificação, armazenamento, recuperação e uso da informação. Estes fluxos foram delineados por Smit e Barreto (2012) e Tarapanoff (2006), sendo aplicáveis aos chamados SIs clássicos.

Ao analisar os SIs em ambientes digitais, observou-se que estes fluxos podem ocorrer de maneira distinta em relação aos chamados SIs clássicos. Portanto, foi igualmente importante considerar não apenas os processos relacionados aos fluxos informacionais nestes ambientes, mas também questões referentes à recuperação da informação. Além disso, foi essencial abordar questões como a precisão da informação recuperada, sua relevância, pertinência e acessibilidade (física e cognitiva), como discutido por Ingwersen (1996), Ingwersen e Jäverlin (2004) e Hjørland (2010).

No contexto deste capítulo, a discussão foi conduzida com base na análise destes aspectos em ambientes digitais. Com isso, foi apresentado como o modelo proposto aprimora estes aspectos para tais ambientes informacionais, indicando as contribuições para a área de CI, que também se preocupa com esse tipo de ambiente informacional.

Tal modelo considerou a utilização dos contextos de uso da informação, especificamente os contextos socioculturais (relativos ao usuário), cognitivos (relativos à necessidade informacional e características da busca) e LSI (relativo as escolhas de pertinência e relevância), a fim de vincular o estoque de ambientes informacionais digitais e as necessidades informacionais de seus potenciais usuários.

Após desenvolvimento e exposição de uma possível aplicação do modelo proposto, foram percebidos pontos contribuintes para SIs nesse tipo de ambiente, bem

como contribuições para a CI. Estes pontos estão compilados na matriz SWOT abaixo (Quadro 7), bem como outros pontos de atenção após análise dos resultados obtidos.

**Quadro 7 - Matriz SWOT para avaliação do modelo proposto**

	<b>FATORES POSITIVOS</b>	<b>FATORES NEGATIVOS</b>
	<b>FORÇAS</b>	<b>FRAQUEZAS</b>
<b>FATORES INTERNOS</b>	<p>Proposição de uma nova estrutura de fluxo informacional para ambientes digitais</p> <p>Aprimoramento do processo de recuperação de informações em ambientes digitais, no que tange relevância, pertinência, precisão e acessibilidade, a partir de proposição de fluxo informacional para este ambiente.</p>	<p>Demanda de implementação para validação da proposição do novo fluxo, bem como da validação e desenvolvimento de aspectos tecnológicos do modelo.</p>
	<b>OPORTUNIDADES</b>	<b>RISCOS (AMEAÇAS)</b>
<b>FATORES EXTERNOS</b>	<p>Mitigar a alta revocação e baixa precisão de digitais, fornecendo respostas mais relevantes e pertinentes para um usuário, através da aplicação conceitual do fluxo informacional proposto.</p>	<p>O desenvolvimento atual dos métodos de TI pode limitar a adaptação do fluxo informacional em ambientes digitais.</p>

Fonte: Elaborado pelo autor.

SWOT é um acrônimo para *Strengths* (Forças), *Weaknesses* (Fraquezas), *Opportunities* (Oportunidades) e *Threats* (Ameaças). Utilizar a análise SWOT na discussão dos resultados da tese confere um arcabouço metodológico robusto para a avaliação crítica do modelo proposto, que visa melhorar a recuperação de informações em ambientes informacionais digitais. Esta técnica, originada na gestão empresarial, é transferível ao contexto científico devido à sua capacidade de isolar e analisar variáveis internas e externas. No interior do modelo, as forças e fraquezas podem ser delineadas, permitindo uma compreensão mais aprofundada de suas

capacidades e limitações inerentes. Externamente, as oportunidades e ameaças apresentam um panorama do ambiente digital e informacional mais amplo, incluindo variáveis como evoluções tecnológicas e padrões de comportamento do usuário. Ao aplicar a análise SWOT, o trabalho adquire uma dimensão analítica mais rica, o que favorece a compreensão dos resultados dentro de uma estrutura conceitual e prática mais ampla, fortalecendo, assim, sua contribuição para o campo da CI.

## **5.2 Discussão**

O modelo foi validado por meio de testes conceituais. Utilizou-se o SRI digital da BRAPCI como referência para essa validação. Durante os testes, o modelo foi aplicado na operação real da BRAPCI, permitindo identificar suas vantagens e desvantagens potenciais em termos de uso e aplicação na CI. A análise SWOT foi adotada para avaliar as características do modelo, oferecendo uma visão abrangente das forças, fraquezas, oportunidades e ameaças, resultando numa análise equilibrada e detalhada para compreensão do modelo proposto no estudo.

Os resultados evidenciaram que a utilização de atributos contextuais aprimorou os processos de recuperação de informações em ambientes informacionais digitais. Quando empregados, tais atributos possibilitam estabelecer uma relação direta entre o perfil do usuário, sua demanda informacional específica e os documentos depositados em repositórios digitais, os quais potencialmente atendem, com precisão e relevância, a referida demanda.

Conforme delineado por Smit e Barreto (2012), Lima e Campos (2022), Novellino (1996) e Ferneda (2003), o paradigma tradicional enfatiza a necessidade de análise prévia dos documentos antes de serem armazenados, visando direcioná-los a um público específico. O foco é a classificação e indexação documental para direcionamento a um usuário específico. Entretanto, nos ambientes digitais contemporâneos, devido à magnitude e diversidade de dados, essa análise prévia muitas vezes é restringida. Em resposta a essa limitação, tem-se dado prioridade ao armazenamento exaustivo de documentos, adiando seu processamento para etapas subsequentes.

Face a tal contexto, propôs-se uma vinculação de informações pós-armazenamento, fundamentada no emprego de atributos contextuais. Nesta abordagem, coletam-se todos os documentos disponíveis em vez de selecioná-los

previamente. Posteriormente, reforçam-se os termos vinculadores nos documentos com base em informações contextuais. Os documentos são classificados e vinculados com base no que os usuários avaliam como importante. Essa abordagem difere das práticas tradicionais, pois se reconhece que, devido a limitações tecnológicas, não é possível analisar todos os documentos detalhadamente desde o início. Em vez disso, a análise acontece depois, usando informações contextuais, como preferências de busca e perfil demográfico, para entender melhor o conteúdo dos documentos. Esse processo é exemplificado na Figura 42 e é fundamentado nas premissas estabelecidas por Lima e Campos (2022), propondo uma reconfiguração dos fluxos informacionais em ambientes digitais.

**Figura 42 - Fluxo informacional do SRI proposto no trabalho, com base na estrutura de Lima e Campos (2022)**



Fonte: Elaborado pelo autor, com base em Lima e Campos (2022).

Ao adotar uma perspectiva posterior no fluxo informacional, o sistema estabelece uma ligação dinâmica entre usuários e documentos, suprimindo os desafios de indexação em ambientes que enfrentam dificuldades em criar indexadores apropriados. Portanto, a reconfiguração do fluxo informacional aprimora a correspondência entre o acervo informacional e o usuário-alvo. Segundo Araújo Júnior (2005), é imperativo direcionar o usuário com precisão para a informação ou documento que atenda à sua demanda intrínseca. Também, o papel do usuário em sistemas de informação é central, conforme pontuado por Frohman (2008) e Ingwersen e Jäverlin (2004).

No caso em análise, quando um estudante de ciências humanas usa o BRAPCI para buscar informações sobre o autor Habermas na área de CI, a mudança no perfil do usuário resultou em diferentes resultados de pesquisa, influenciados por vários fatores contextuais. Dessa forma, o sistema mostrou a capacidade de organizar as informações após o armazenamento, eliminando a necessidade de análises prévias para determinar o que deve ser mantido.

Em complemento, o modelo introduzido visou otimizar a recuperação de informações por meio de um mecanismo de avaliação constante. Em sua aplicação, sempre que um usuário interage com o sistema e recupera informações, o referido usuário tem a oportunidade de avaliar a relevância do conteúdo fornecido. Se uma informação não atender às suas expectativas, ele tem a opção de rejeitá-la, forçando assim o sistema a adaptar-se e refinar suas buscas. Assim, graças à habilidade do modelo de se adaptar às mudanças no perfil ao longo do tempo e à possibilidade de o usuário avaliar a informação que recebe, o sistema pôde sintonizar-se com o que cada usuário específico realmente precisa. Tal combinação de ajuste de perfil e avaliação subsequente das informações permitiu que o sistema se adaptasse de acordo com as particularidades de cada perfil de usuário.

Esta abordagem foi fundamentada em análise de teorias sobre relevância informacional, apoiando-se nas contribuições teóricas de Choo (2003) e Campbell e Van Rijsbergen (1996). Estes teóricos sublinham a essência subjetiva do usuário e a importância do contexto no uso da informação. Complementarmente, as abordagens de Mizzaro (1998) e Reid (1999) são integradas para garantir que a relevância seja examinada tanto sob a ótica do contexto de uso, quanto da interpretação individual da informação. Este modelo de análise de relevância direciona a avaliação para o contexto de uso efetivo da informação, priorizando a necessidade real de informação do usuário. Tal abordagem foi complementada por Reid (1999), o qual denominou esse tipo de relevância de "relevância da tarefa" ou relevância situacional, com o objetivo de capturar a utilidade percebida pelo usuário e os objetos de informação recuperados em relação à tarefa e à necessidade real de informação. Esse conceito de relevância também considera as necessidades de uso da informação, sejam estas objetivas ou subjetivas.

Sobre a utilização de atributos contextuais, muito embora seja viável desenvolver uma lista inicial de atributos contextuais com base na literatura e teorias existentes, a definição específica e taxonômica de tais atributos requer uma

abordagem mais aprofundada, incluindo testes conceituais em cenários de recuperação de informação.

Tais testes e experimentos são necessários para avaliar a relevância e o impacto de cada atributo no processo de recuperação de informação. Esses testes empíricos exigem a implementação em linguagem de programação, mantendo a abordagem de enquadramento proposta por Johnson (2003) para a estruturação de atributos contextuais em sistemas sensíveis ao contexto.

Além disso, Henrique, Nassif e Venâncio (2007) expandiram a proposta de Johnson (2003), destacando a importância crescente de identificar as ações dos indivíduos em seu domínio de atuação ou contexto de observação. Isso envolve entender os contatos que eles estabelecem no dia a dia, suas predisposições, interesses relacionados ao domínio de atuação e sua disposição emocional e afetiva em detalhes.

Tal problemática diz respeito ao desafio da incapacidade lógica de identificação de atributos contextuais. A incapacidade lógica pode ser entendida como a dificuldade de traduzir de forma efetiva os atributos contextuais capturados em variáveis que possam ser usadas na linguagem de programação para operar o modelo. A complexidade e variedade dos atributos socioculturais tornam desafiador criar um algoritmo que possa abrangê-los de forma completa e adequada. Também, é importante abordar o problema de informações falsas e imprecisas que podem ser coletadas como atributos contextuais, exigindo estratégias para mitigar esse problema.

Além disso, a interpretação e significado dos atributos contextuais podem variar entre os usuários, tornando ainda mais difícil sua codificação em um algoritmo geral. Isto é particularmente relevante ao capturar atributos cognitivos, onde conjuntos de termos semelhantes podem ter significados diferentes. Portanto, realizar treinamentos e testes práticos dos instrumentos de descoberta de conhecimento em textos é necessário para garantir que as combinações contextuais sejam tratadas adequadamente.

Além disso, existe o desafio da capacidade de processamento computacional atual para capturar e processar atributos contextuais em grande escala, apesar dos avanços nas tecnologias de processamento de linguagem natural, como o GPT. Isso se torna especialmente problemático ao lidar com grandes volumes de dados produzidos rapidamente, pois as tecnologias de processamento de dados podem

enfrentar restrições de tempo de execução que limitam sua capacidade de análise documental abrangente.

É fundamental também considerar o risco de um ajuste excessivo nas relações entre mapas contextuais e documentos, o que pode prejudicar a relevância da informação recuperada, como observado por Savolainen e Thomson (2022). Portanto, encontrar um equilíbrio entre a personalização e a diversidade de informações relevantes é essencial, especialmente considerando as limitações tecnológicas no processamento de dados.

O modelo proposto, não considera inicialmente estes fatores, tendo em vista a demanda por testes empíricos com o modelo, após implementação em linguagem de programação.

Apesar dos desafios mencionados, o modelo conceitual proposto oferece uma oportunidade de aprimorar os SIs em ambientes digitais. A aplicação desse modelo visa melhorar a precisão das informações recuperadas em bases de dados digitais, que frequentemente contêm uma grande variedade de informações em diferentes formatos e fontes, como mencionado pelos autores e por Johnson (2003). Ao considerar o contexto e os atributos relevantes, o modelo permite uma recuperação de informações mais personalizada, melhorando a experiência do usuário e aumentando a relevância dos resultados apresentados.

A integração desse fluxo em ambientes digitais é postulada, não somente como uma mudança metodológica, mas também como uma contribuição à CI. Esta área, corroborado pelas reflexões de Tarapanoff (2006), permanece em contínua busca por estratégias otimizadas para enfrentar os desafios atrelados ao acesso informacional em plataformas digitais. O fluxo informacional proposto demonstrou sua eficácia ao recuperar informações de com relevância e pertinência.

## **CAPÍTULO 6 – CONCLUSÕES E CONSIDERAÇÕES FINAIS**

Este estudo teve como objetivo a proposição de um modelo de recuperação de informações destinado a aprimorar o acesso à informação em ambientes digitais de informação. Tal necessidade se fundamenta na ausência de instrumentos que permitam uma identificação completa do usuário nesses ambientes, resultando em uma recuperação de informações deficitária, incapaz de atender adequadamente as demandas informacionais dos usuários. A hipótese subjacente a este trabalho é que, dado que o uso da informação por parte do usuário pode ocorrer em contextos diversos, dependentes de seu perfil e necessidades informacionais específicas, o contexto de uso da informação pode ser empregado como um elemento que aprimora o processo de recuperação de informações em tais ambientes digitais.

Para a concepção deste modelo, foram identificadas, na perspectiva da CI, as características dos ambientes informacionais digitais contemporâneos. Adicionalmente, foram identificadas variáveis intrínsecas ao contexto de uso, que podem servir como elementos de apoio no desenvolvimento do referido modelo.

Conforme apontado por Davenport (2014) e Marquesone (2016), devido às restrições de processamento e computação impostas pelos algoritmos computacionais nos ambientes de TDIC em relação aos SIs tradicionais, tornou-se imperativo investigar a interação entre essas duas áreas. Este estudo visou aprofundar a compreensão sobre como um modelo poderia enfrentar os desafios apresentados por esses ambientes digitais, com base nas teorias da CI, especialmente aquelas relacionadas ao contexto e aos fluxos de informação nos SIs.

Portanto, não foi possível estudar os SIs digitais sem considerar cuidadosamente suas tecnologias computacionais. Tal abordagem contribuiu significativamente para o campo da CI, proporcionando uma compreensão mais abrangente dos obstáculos e do acesso à informação em tais ambientes digitais. Como resultado deste estudo, desenvolveu-se um modelo que demonstrou aprimoramento na recuperação de informações relevantes, elevando, assim, a qualidade dos fluxos de informação por meio da exploração dos contextos de utilização da informação em ambientes digitais.

Também, no contexto deste trabalho, foi imprescindível sublinhar a importância da interdisciplinaridade nas abordagens metodológicas e teóricas adotadas. Em especial, este trabalho se apoiou na Teoria da Ação proposta por Donald Norman (1986) e na norma NBR 9241-11. Estas referências são particularmente relevantes por serem tradicionalmente consideradas como pontos de partida no desenvolvimento

e compreensão de sistemas digitais, bem como no entendimento do usuário e seu contexto de ação.

É importante destacar que a norma NBR 9241-11 engloba aspectos ergonômicos de dispositivos integrantes de sistemas de uso mais abrangentes, mas também se aplica na concepção de sistemas de forma geral. Isso é especialmente relevante para sistemas que envolvem a interação de usuários e seus respectivos contextos. De forma complementar, a Teoria da Ação de Donald Norman (1986) ofereceu um quadro analítico que auxilia na compreensão da formação e contextualização das ações e decisões dos usuários dentro desses sistemas unificados.

Este alinhamento entre a Teoria da Ação e a norma NBR 9241-11 com os conceitos de sujeito informacional e contexto, conforme apontou Gonzalez de Gómez (2017), teve como objetivo a criação de um modelo teórico e metodológico. Neste modelo, considerou-se o usuário como um ator social cujas necessidades e ações informacionais estão profundamente entrelaçadas com o contexto tecnológico e social em que estão inseridas.

Ao combinar estas diferentes perspectivas, a tese buscou estabelecer um vínculo entre as intrínsecas demandas do usuário, como delineadas pela norma, e suas ações e necessidades em um ambiente de sistemas digitais. Este enfoque multidisciplinar possibilitou que o modelo desenvolvido fosse adaptável a uma variedade de contextos e necessidades informacionais, conforme observado em sua aplicação na BRAPCI, reforçando a relevância do contexto no campo da CI e contribuindo para uma compreensão mais completa e adaptável do sujeito informacional.

Portanto, o modelo proposto de recuperação de informação desempenhou um papel significativo na melhoria da relevância e pertinência da recuperação de informações em ambientes informacionais digitais. Ao considerar um fluxo informacional específico para estes ambientes, o modelo conseguiu harmonizar as ferramentas tecnológicas de TI com as preocupações fundamentais da CI. Este fluxo informacional adaptado levou em consideração a relação entre o estoque de informações e o usuário, a organização eficiente desse estoque e a disponibilidade de instrumentos de recuperação relevantes e pertinentes para acesso à informação. Como resultado, o modelo contribuiu para aprimorar a qualidade dos fluxos informacionais nos referidos ambientes, tornando-os mais adequados às demandas

dos usuários e oferecendo informações cada vez mais relevantes e úteis. Para tanto, o modelo proposto utilizou atributos contextuais de uso da informação, vinculando peculiaridades de busca dos usuários a documentos pertinentes.

Reconhece-se, entretanto, os desafios: a dificuldade de incorporar atributos contextuais nas tecnologias de informação, a influência da velocidade de atualização e o risco de ajuste excessivo do modelo, podendo resultar em informações enviesadas ou limitadas.

## **6.1 Conclusões**

No que diz respeito ao objetivo geral deste trabalho, no desenvolvimento do modelo de recuperação de informações, a pesquisa deparou-se com uma série de desafios inerentes à recuperação de informações em ambientes digitais. Tais desafios abrangeram desde a identificação potencial dos usuários do sistema até a análise e prévia identificação de documentos, passando pela complexa tarefa de indexação e classificação dos documentos a serem armazenados. Além disso, o estudo confrontou-se com a necessidade de lidar com grandes volumes de dados em alta velocidade, o que acentuou a variabilidade dessas informações. Somando-se a isso, as dificuldades na vinculação desses documentos e arquivos armazenados constituíram desafios significativos no contexto da pesquisa.

Devido à constante produção e grande quantidade de informações nestas bases, as técnicas tradicionais de preparação documental podem causar gargalos no sistema. Novos regimes informacionais, impulsionados pelas tecnologias digitais e móveis, levaram a uma produção desestruturada e em grande quantidade de documentos e arquivos, como observado por Gonzalez de Gómez (2012) criando desafios na seleção, classificação, representação e indexação dessas informações armazenadas.

Para mitigar esses desafios, são normalmente utilizados métodos e tecnologias, como bancos de dados NoSQL e algoritmos de aprendizado de máquina, para organizar e vincular os dados para consultas futuras (DAVENPORT, 2014; WHITE, 2015). A partir destas tecnologias, todas as informações produzidas são armazenadas, e processadas de forma posterior.

No entanto, estes processos ainda apresentam limitações, resultando em altos índices de revocação de informações e baixa especificidade dos metadados extraídos

dos documentos. A geração de metadados para vinculação entre documentos e consultas muitas vezes considera a exaustividade máxima possível de termos em um documento, o que pode resultar na inclusão de termos irrelevantes para determinados contextos de uso da informação, prejudicando a correspondência entre as consultas futuras e as necessidades informacionais dos usuários.

Mesmo com o uso de técnicas avançadas, como o GPT, que utiliza redes neurais de atenção (VASWANI *et al.*, 2017), estas abordagens podem enfrentar limitações ao lidar com o grande volume e a diversidade de informações provenientes de diferentes contextos em ambientes informacionais digitais. Observa-se que o problema de vinculação entre consulta e busca devido à estrutura de organização e indexação dos documentos e arquivos, neste tipo de ambiente, é patente. Neste sentido, uma representação deficiente da informação pode levar a uma recuperação de informação insatisfatória.

Este estudo sugere uma abordagem que prioriza a coleta de todos os documentos disponíveis e, posteriormente, reforça os termos vinculadores com base em informações contextuais, visando à organização e conexão dos documentos de acordo com a importância avaliada pelos usuários. Esta reconfiguração, ilustrada na figura 41, visa superar desafios de armazenamento e processamento. Mesmo com técnicas avançadas, como o GPT que utiliza redes neurais de atenção (VASWANI *et al.*, 2017), a vinculação entre consulta e busca pode ser comprometida. Assim, a proposta busca alinhar as necessidades dos usuários com as informações recuperadas, melhorando a precisão da recuperação de informação.

Sobre os objetivos específicos, conclui-se que:

(1) A operação de sistemas de recuperação de informação em ambientes digitais enfrenta um desafio crítico relacionado à não identificação dos usuários. A dificuldade reside em direcionar efetivamente as buscas quando os usuários permanecem anônimos, o que limita a capacidade de personalizar os resultados de acordo com suas necessidades e contextos individuais. Tal ausência de identificação dificulta a entrega de informações qualificadas e relevantes, representando um dos principais obstáculos na operação desses sistemas (BARLOW, 2013).

Ambientes informacionais digitais lidam com a manipulação de grandes volumes de dados, com alta velocidade e variabilidade midiática. Estas características tornam complexa a identificação seus potenciais usuários que podem buscar informações nestes ambientes (DAVENPORT, 2014), tendo em vista que esta

produção pode considerar as mais variadas temáticas, bem como os mais variados públicos-alvo (BRAQUEHAIS; MORESI; WILBERT, 2017). Sabe-se que sistemas de informação devem ser projetados com um usuário em vista, e que a não adoção dessa estratégia impacta negativamente os processos de recuperação de informação (FROHMAN, 2008).

A natureza dos ambientes digitais em questão afeta a recuperação de informações relevantes. Nestas circunstâncias, devido à falta de análise prévia do que deve ser armazenado, todo o conteúdo produzido é considerado para armazenamento. Consequentemente, a recuperação de informações é realizada por meio de técnicas baseadas em algoritmos estatísticos. Isto implica que as informações são recuperadas não para atender a uma demanda informacional específica, mas com o potencial de atender a diversas demandas potenciais que possam se beneficiar das informações recuperadas.

Isso converge com o apresentado por Smit e Barreto (2002) e Tarapanoff (2006) sobre fluxos informacionais em SIs. A fim de recuperar a melhor informação possível para um usuário, é necessária análise documental para verificar se um determinado conjunto de documentos será considerado para armazenamento. Após este processo, ele é preparado e indexado em relação à um perfil de usuário específico, e sua potencial necessidade (FERNEDA, 2003, LIMA; CAMPOS, 2022). De forma geral, o processo observado em ambientes digitais não respeita este fluxo informacional.

No contexto descrito, tornou-se essencial o desenvolvimento de estratégias e tecnologias adequadas para lidar com a identificação dos usuários e a recuperação de informações relevantes em tais ambientes. Devido à natureza transdisciplinar do estudo destes ambientes, foi necessário repensar métodos e técnicas que relacionassem a TI e a CI, a fim de encontrar meios tecnológicos que estivessem alinhados com os princípios e objetivos da CI. Assim, a presente pesquisa considerou a incorporação de atributos contextuais e personalização dos sistemas de recuperação de informação, de forma a adequar as respostas às necessidades específicas de cada usuário.

(2) Na revisão realizada, identificaram-se três campos de atenção contextual relevantes para a RI: sociocultural, linguístico/simbólico/interpretativo e cognitivo. Cada um destes campos possui variáveis e condições específicas que afetam o

contexto de uso da informação em sistemas de informação e recuperação da informação, sendo importantes para melhorar os SIs em ambientes digitais.

No campo sociocultural, as variáveis incluem a formação do sujeito, interações sociais, meio ambiente e aspectos históricos e culturais. Estes elementos influenciam a relação do usuário com a informação e suas demandas informacionais, pois cada indivíduo possui uma bagagem sociocultural única que impacta suas necessidades e preferências na recuperação da informação.

O campo linguístico/simbólico/interpretativo engloba a percepção do mundo e a comunicação do usuário. Neste campo, estão presentes aspectos estéticos, subjetivos e linguísticos, que influenciam na seleção e apresentação da informação. A forma como o usuário percebe e interpreta o mundo ao seu redor, bem como sua comunicação através de linguagem, jargões e regionalismos, são variáveis importantes a serem consideradas na recuperação da informação.

Esta perspectiva do campo linguístico/simbólico/interpretativo complementa as abordagens de Choo (2003), Campbell e Van Rijsbergen (1996), Mizzaro (1998) e Reid (1999) mencionadas anteriormente. Estes autores destacam a importância de considerar as questões de situação, emoção, interpretação e relevância da informação para o usuário.

Ao levar em conta o contexto de uso da informação, a subjetividade do usuário na avaliação da relevância, a seleção dos documentos adequados e a utilidade percebida em relação à tarefa e à necessidade real de informação, o modelo proposto busca abarcar a complexidade e a subjetividade inerentes ao processo de recuperação da informação. Assim, é possível oferecer uma abordagem mais abrangente e precisa na entrega de informações relevantes aos usuários. Já o campo cognitivo está relacionado à linguagem e ao conhecimento prévio do usuário. Aqui, são considerados fatores mais lógicos do que estéticos, envolvendo a compreensão e o processamento da informação. Elementos como idioma, terminologia específica e conhecimento prévio do usuário desempenham um papel fundamental na recuperação da informação e na apresentação de resultados adequados.

Estes três campos de atenção contextual forneceram uma visão abrangente dos elementos que influenciam o contexto de uso da informação em sistemas de recuperação da informação. Ao trabalhar estes elementos, foi possível apresentar um modelo conceitual aprimorado, adaptando a apresentação da informação de acordo com as características socioculturais, linguísticas e cognitivas dos usuários. Isto

resultou em uma recuperação de informação mais eficiente e relevante, atendendo às necessidades informacionais específicas de cada usuário.

No SRI proposto, o usuário descreve seu perfil e explicita seu problema ou necessidade informacional. O sistema identifica termos e repetições de palavras na busca, compondo os atributos contextuais cognitivos. Estes atributos, juntamente com os atributos contextuais socioculturais do usuário, são armazenados em um BD provisório utilizando XML. No primeiro ciclo de busca, o sistema recupera documentos e arquivos que contenham conjuntos de palavras similares aos termos de busca.

O usuário avalia a relevância das informações recuperadas, e documentos relevantes ou pertinentes têm seus termos vinculadores atualizados com os dados contextuais do usuário. Isso refina a forma como os documentos são relacionados aos usuários, considerando seu perfil e sua necessidade. O objetivo, assim, foi de recuperar informações com maior precisão, levando em conta os atributos contextuais. A partir desta operação, o sistema demonstrou efetividade na recuperação de informações relevantes, sustentado por uma abordagem embasada em teorias da área da CI, em especial àquelas envolvendo o estudo de fluxos informacionais e sistemas da informação.

Em conclusão, esta tese apresenta um SRI que leva em consideração atributos contextuais do usuário, fundamentado em teorias da CI. O sistema demonstrou eficácia na recuperação de informações relevantes, indo além da indexação baseada em palavras-chave ao incorporar aspectos contextuais.

A relevância do algoritmo desenvolvido se estende à BRAPCI, objeto empírico em que o modelo foi analisado. Embora inicialmente destinada a pesquisadores da área de CI, a BRAPCI também é útil para acadêmicos de outras disciplinas. O algoritmo proposto aprimora essa utilidade ao oferecer indexação e acessibilidade mais diversificadas. Essa abordagem é apoiada pelo trabalho de Gonzalez de Gómez (2017), que destaca a importância do contexto no comportamento informacional.

Além disso, o algoritmo tem o potencial de expandir a aplicabilidade da BRAPCI sem alterar sua infraestrutura existente. O sistema não só mantém sua utilidade para os pesquisadores da área de CI como também a torna mais acessível para acadêmicos de diferentes campos. Isso é conseguido por meio da inclusão de diferentes níveis de acessibilidade, tornando a BRAPCI mais adaptável a diferentes necessidades informacionais. Assim, a escolha da BRAPCI como objeto de estudo nesta tese é justificada pela sua utilidade existente e pelo potencial de aprimoramento

que o algoritmo proposto oferece, tanto para a CI quanto para outros campos acadêmicos.

## **6.2 Limitações da pesquisa**

Em termos de limitações, primeiramente, é importante ressaltar que o modelo proposto ainda precisa ser testado exaustivamente e validado em outros cenários além do contexto em que foi desenvolvido. Essa ampliação de aplicação permitiria avaliar a eficácia do modelo em diferentes contextos de uso da informação.

Além disso, é necessário levar em conta que a implementação do modelo está sujeita às tecnologias disponíveis atualmente. Isso significa que o desempenho e os resultados obtidos podem ser influenciados pelas limitações tecnológicas existentes. Portanto, é importante considerar futuros avanços tecnológicos que possam contribuir para aprimorar o modelo e superar essas restrições.

Outra limitação a ser mencionada é a necessidade de uma validação mais abrangente do modelo. Embora conceitualmente o modelo apresente fundamentos sólidos, é necessário realizar uma validação mais exaustiva, incluindo testes práticos e coleta de dados reais dos usuários. Essa validação contribuiria para confirmar a eficácia e a utilidade do modelo em situações reais de uso.

Esta pesquisa propõe alternativas no campo da CI para a operação desses ambientes digitais, explorando o potencial das ferramentas de TDIC. Tais soluções permitem superar desafios na gestão e acesso à informação em ambientes digitais. No entanto, é importante reconhecer que esta abordagem difere dos fluxos informacionais tradicionais discutidos por Smit e Barreto (2012) e Tarapanoff (2006).

Enquanto tais fluxos priorizam a análise documental prévia e classificação inicial, o modelo proposto se baseia na captura de atributos contextuais e na vinculação posterior entre documentos e usuários. Esta abordagem confia fortemente em algoritmos para os processos de vinculação documental, sem a intervenção direta de um agente cognoscível. Portanto, o fluxo proposto não possui um "grau de controle observável" em sua ação, o que demanda estudos futuros sobre como operar e lidar com essas vinculações à luz da CI.

Com o passar do tempo, a operação do algoritmo utilizado pode tornar o comportamento desse fluxo proposto imprevisível. Portanto, como parte do processo

de consolidação, são necessários estudos de acompanhamento para mitigar possíveis problemas, levando em consideração as preocupações relacionadas à relação entre informações e usuários, assim como à recuperação de informações relevantes nesses ambientes.

Também, no presente trabalho, a coleta de atributos contextuais socioculturais dos usuários é realizada através de formulários. Esta estratégia, contudo, enfrenta várias limitações. Primeiramente, usuários podem se recusar a preencher tais formulários, o que compromete a extensão e a qualidade dos dados coletados. Além disso, o fornecimento de informações imprecisas ou incorretas representa outro ponto de preocupação, questionando a confiabilidade dos dados obtidos.

No contexto brasileiro, é imperativo que essa abordagem de coleta de dados esteja em conformidade com o Marco Civil da Internet (Brasil, Lei nº 12.965, de 23 de abril de 2014). Este marco legal estabelece princípios, garantias, direitos e deveres para o uso da internet no Brasil. Entre seus principais pontos, destaca-se a necessidade de obtenção do consentimento expresso do usuário para a coleta de dados pessoais, bem como a garantia da privacidade e da segurança desses dados durante seu armazenamento e processamento. Portanto, qualquer metodologia de coleta de dados deve ser estruturada de forma a respeitar esses princípios legais, garantindo a integridade e a confidencialidade das informações coletadas.

Por fim, é importante considerar as restrições do equipamento utilizado na pesquisa. Dependendo dos recursos disponíveis, pode haver limitações em relação à capacidade de processamento, armazenamento e velocidade de resposta do sistema. Estas restrições podem afetar o desempenho do modelo e, portanto, devem ser levadas em conta ao avaliar os resultados obtidos.

Em suma, embora o modelo proposto apresente um embasamento conceitual sólido, é necessário considerar as limitações relacionadas à sua aplicação em outros cenários, a dependência das tecnologias atuais, a necessidade de validação abrangente e as restrições do equipamento utilizado. Estas limitações destacam a importância de futuros estudos e aprimoramentos para aperfeiçoar o modelo e expandir sua aplicabilidade.

### 6.3 Trabalhos futuros

Diversas oportunidades de pesquisa e aprimoramento surgem a partir das limitações identificadas nesta tese. Um caminho promissor é a implementação do modelo, visando melhorar a precisão das informações recuperadas. Embora o modelo conceitual esteja fundamentado em sólidos alicerces teóricos, sua efetiva aplicação e validação requerem a implementação prática em um ambiente real. Portanto, é recomendado o desenvolvimento de protótipos ou sistemas funcionais que permitam testar e validar o modelo em cenários reais de uso. Essa operacionalização proporcionaria dados valiosos sobre o desempenho, a usabilidade e a eficácia do modelo, além de permitir a coleta de dados empíricos para sua validação.

Sobre usuários nesses ambientes informacionais digitais, pesquisas alicerçadas nas bases teóricas da CI podem contribuir para aprimorar a interação entre os usuários e os sistemas utilizados nestes ambientes. Investigar a experiência do usuário e a usabilidade de sistemas e interfaces nos fluxos informacionais de ambientes digitais permitiria identificar os desafios enfrentados pelos usuários na busca, seleção e recuperação de informações. Com base nessas informações, seria possível propor melhorias no projeto de sistemas e interfaces, tornando os processos informacionais mais intuitivos e eficazes.

Ainda sobre estudos de usuário, pode ser útil a investigação da utilização de contextos de uso para identificação do perfil do usuário de forma prévia nos fluxos informacionais. Essa pesquisa buscaria explorar métodos e técnicas que permitam identificar e compreender o contexto de uso da informação antes mesmo da interação direta do usuário com o sistema. A partir dessas informações, seria possível criar perfis de usuários mais precisos e personalizados, permitindo uma adaptação mais eficiente dos fluxos informacionais às necessidades individuais de cada usuário.

Isso poderia resultar em uma experiência de informação mais satisfatória, com recomendações mais relevantes, sugestões personalizadas e maior eficiência na recuperação de informações. Para tanto, esta pesquisa poderia explorar técnicas de análise de dados, mineração de dados, IA e aprendizado de máquina para identificar padrões e correlações entre o contexto de uso e as preferências do usuário. Além disso, questões éticas relacionadas à privacidade e segurança dos dados dos usuários também seriam consideradas e abordadas neste estudo.

Também, os trabalhos futuros devem priorizar a realização de testes e experimentos empíricos para avaliar a relevância e o impacto dos atributos contextuais na recuperação da informação. Além disso, é recomendado explorar ambientes físicos para conduzir experimentos práticos e coletar dados reais dos usuários, a fim de obter dados valiosos sobre a eficácia e a utilidade dos atributos em diferentes cenários de busca e recuperação de informações. Estas iniciativas podem contribuir para um aprimoramento contínuo dos sistemas sensíveis ao contexto e para uma melhor compreensão da influência dos atributos contextuais na recuperação da informação.

O desenvolvimento de métodos de classificação e organização automática de informações em ambientes digitais poderia dar cabo dessa questão. Investigar e desenvolver métodos automatizados que possam classificar e organizar informações de forma prévia, com base no apontado por Smit e Barreto (2012) e Tarapanoff (2006), por exemplo, considerando as características específicas desses ambientes, seria um objetivo relevante. Esta pesquisa poderia explorar técnicas de aprendizado de máquina e processamento de linguagem natural para facilitar a classificação e recuperação eficiente de informações de forma prévia, garantindo estoques mais direcionados a partir de um grande fluxo de informações.

Outro trabalho futuro e necessário é a definição de estratégias para promover a acessibilidade informacional, utilizando a captação de contextos de uso da informação como ferramenta chave neste processo. Conforme ressaltado por Fujino (2017),

[...] na Ciência da Informação, a acessibilidade informacional depende do conhecimento das necessidades de informação do potencial usuário, além dos aspectos que envolvem a infraestrutura para acesso e divulgação, condições fundamentais para o desenvolvimento de mediações que viabilizem o acesso cognitivo e apropriação das informações pelo usuário (FUJINO, 2017, p. 237).

Para avançar neste campo, será fundamental realizar estudos e pesquisas que explorem como a captação de contextos de uso da informação pode ser aplicada de forma efetiva na promoção da acessibilidade informacional. Este trabalho futuro possibilitaria desenvolver estratégias mais direcionadas e personalizadas, levando em consideração as características individuais dos usuários e proporcionando uma experiência de acesso à informação mais inclusiva e eficaz.

Isso posto, pesquisa futura também deverá se dedicar à exploração de métodos alternativos para a coleta eficiente e ética de atributos contextuais, respeitando as diretrizes legais e contribuindo para uma compreensão mais rica e multidimensional do sujeito informacional e de seu contexto. Esta investigação subsequente não somente tem o potencial de aprimorar o modelo proposto nesta tese, como também de oferecer percepções valiosas para a área da CI como um todo.

Assim, os trabalhos futuros para continuidade desta pesquisa devem se concentrar no aprimoramento de sistemas de informação digitais, no desenvolvimento de técnicas avançadas de aprendizado de máquina e processamento de linguagem natural para a captura de atributos contextuais, na identificação detalhada dos atributos contextuais necessários para aprimoramento de recuperação da informação, no desenvolvimento de ambientes acessíveis e na operacionalização do modelo proposto para sua validação. Estas iniciativas podem contribuir para o desenvolvimento de sistemas com recuperação de informações mais precisa e personalizada, melhorando a experiência dos usuários e impulsionando o avanço da pesquisa no campo de SIs em ambientes digitais.

## Referências<sup>22</sup>

ABBASI, A. **Learning Apache Spark 2**. Nova Iorque: Packt, 2017.

ABNT. **NBR 9241-11** : Requisitos Ergonômicos para Trabalho de Escritórios com Computadores Parte 11 – Orientações sobre Usabilidade. Rio de Janeiro: ABNT, 2011.

ABOWD, G.; DEY, A. Towards a Better Understanding of Context and Context-Awareness. Lecture Notes in Computer Science. **Anais** [...]. Atlanta: Graphics, Visualization and Usability Center and College of Computing, p. 1-12, jan. 1999.

ADAMS, D. **O Guia do Mochileiro das Galáxias**. São Paulo: Arqueiro, 2004.

ADAMS, N.; SCHILIT, B.; WANT, R. Context-Aware Computing Applications. Workshop on Mobile Computing Systems and Applications. *In*: First workshop on mobile computing systems and applications. **Anais** [...]. Santa Cruz: IEEE Computer Society Press, p. 85-90, dez. 1994.

ANGELONI, M. T. Elementos intervenientes na tomada de decisão. **Ciência da Informação**, v. 1, n. 32, p. 17-22, 2003.

ARAÚJO, C. A. Á. Práticas informacionais: novo conceito para o estudo dos usuários da informação. *In*: **Informação: agentes e intermediação**. Brasília: IBICT, 2017, p. 195-235.

ARAÚJO, V. M. R. H. DE. Sistemas de informação: nova abordagem teórico-conceitual. **Ciência da Informação**, v. 24, n. 1, p. 1-39, 1995.

ARAÚJO JÚNIOR, R. H. **Precisão no Processo de Busca e Recuperação da Informação**. 2005. Tese. - Universidade de Brasília, Brasília, 2005.

ARTERO, A. O. **Inteligência Artificial**. São Paulo: Livraria da Física, 2012.

ASSIS, E.; GOUVÊA JR., M. Aprendizagem por Reforço com Rede Neural no Desenvolvimento de Jogos Digitais. **Instituto de Ciências Exatas e Informática**, v. 1, n. 1, p. 1-5, maio 2017.

AWAD, E.; GHAZIRI, H. **Knowledge Management**. Patparganj: Dorling Kindersley, 2007.

BALCI, O. Verification, Validation and Testing Technics. *In*: **Handbook of Simulation**. Georgia: Jerry Banks, 1998, p. 68 - 83.

BALCI, O. Conceptual modelling for designing large-scale simulations. **Journal of Simulation**, v. 1, n. 1, p. 175-186, 2007.

---

<sup>22</sup> De acordo com a Associação Brasileira de Normas Técnicas (ABNT NBR 6023).

BARBOSA, S. D. J.; SILVA, B. S. **Interação Humano-Computador**. Rio de Janeiro: Elsevier Science, 2010.

BARLOW, M. **The Culture of Big Data**. Sebastopol: O'Reilly Media, 2013.

BARRETO, A. A questão da informação. **São Paulo em Perspectiva**, v. 8, n. 4, p. 3-8, 1994.

BARUZZO, A. et al. A conceptual Model for Digital Libraries Evolution. **Proceedings of the Fifth International Conference on Web Information Systems and Technologies**, p. 299–304, 2009.

BAUMAN, Z. **Modernidade Líquida**. São Paulo: Schwarcz, 2001.

BAUMGÄRTNER, T.; RIBEIRO, F. R.; REIMERS, N.; GUREVYCH, I. **Incorporating Relevance Feedback for Information-Seeking Retrieval using Few-Shot Document Re-Ranking**. Ubiquitous Knowledge Processing Lab (UKP Lab), Department of Computer Science and Hessian Center for AI (hessian.AI), Technical University of Darmstadt, 2022. Disponível em: <arXiv:2210.10695v1 [cs.LG]>. Acesso em: 31/08/2023.

BELKIN, N. Anomalous states of knowledge as a basis for information retrieval. **Canadian Journal of Information Science**, v. 5, n. 1, p. 133–143, 1980.

BIO, S. R. **Sistemas de Informação: um enfoque gerencial**. São Paulo: Atlas, 1996.

BITTENCOURT, G. **Inteligência artificial: Ferramentas e teorias**. Florianópolis: Editora da UFSC, 2006.

BORKO, H. Information Science: What is it? **American Documentation**, v. 9, n. 1, p. 3–5, 1968.

BORLUND, P. The Concept of Relevance in IR. **Journal of the Association for Information Science and Technology**, v. 54, n. 10, p. 913–925, 2003.

BOYD, D.; CRAWFORD, K. Critical Questions for *Big Data*: Provocations for a cultural, technological, and scholarly phenomenon. **Information, Communication & Society**, v. 15, n. 5, p. 662–679, maio 2012.

BRAGA, A.; CARVALHO, A.; LUDERMIR, T. **Redes Neurais Artificiais: Teorias e Aplicações**. Rio de Janeiro: LTC, 2000.

BRAQUEHAIS, A.; MORESI, E.; WILBERT, J. Fatores críticos de sucesso de gestão do conhecimento aplicáveis ao Big Data. Comunicação Oral apresentado em XVIII Encontro Nacional de Pesquisa em Ciência da Informação – ENANCIB 2017. **Anais [...]**. Marília, 2017, p. 1-20.

BRASIL. **Lei nº 12.965, de 23 de abril de 2014**. Estabelece princípios, garantias, direitos e deveres para o uso da Internet no Brasil. Diário Oficial da União, Brasília, DF, 24 abr. 2014. Disponível em: <https://www.gov.br/legislacao/leis/2014/lei-12965-de-23-de-abril-de-2014>. Acesso em: 10 set. 2023.

BUCKLAND, M. Information as a thing. **Journal of the American Society for Information Science**, v. 45, n. 5, p. 351–360, jun. 1991.

BUFREM, L. S. et al. Modelizando práticas para a socialização de informações: a construção de saberes no ensino superior. **Perspectivas em Ciência da Informação**, v. 15, n. 2, p. 22–41, 2010.

CAMPBELL, I.; VAN RIJSBERGEN, K. The ostensive model of developing information needs. *In*: P. **Proceedings of CoLIS 2, second international conference on conceptions of library and information science**: Integration in perspective. Copenhagen: Royal School of Librarianship, 1996. p. 251–268.

CAPURRO, R.; HJØRLAND, B. The concept of information. **Annual Review of Information Science and Technology**, v. 37, n. 1, p. 343–411, ago. 2003.

CESARINO, M. A. DE N. Bibliotecas especializadas, Centros de Documentação, Centro de Análise da Informação: apenas uma questão de terminologia ? **Revista da Escola de Biblioteconomia da UFMG**, v. 7, n. 2, p. 218–241, 1978.

CHARNIAK, E.; MCDERMOTT, D. **Introduction to Artificial Intelligence**. Boston: Addison-Wesley, 1985.

CHOO, C. W. Como ficamos sabendo – um modelo de uso da informação. *In*: **A organização do Conhecimento**: como as organizações usam a informação para criar significado, construir conhecimento e tomar decisões. São Paulo: Editora SENAC, 2003. p. 61–120.

COEIRA, E. W.; VICKLAND, V. Is relevance relevant? User relevance ratings may not predict the impact of Internet search on decision out- comes. **Journal of the American Medical Informatics Association**, v. 15, n. 4, p. 542–545, 2008.

COHEN, D. M. **O consumidor da informação documentária: o usuário de sistemas documentários visto sob a lente da análise documentária**. Tese – Universidade de São Paulo, São Paulo, 1995.

COOL, C.; SPINK, A. Issues of context in information retrieval (IR): an introduction to the special issue. **Information Processing and Management**, v. 38, n. 1, p. 605–611, 2002.

COPPIN, B. **Inteligência Artificial**. Rio de Janeiro: LTC, 2010.

CORMEN, T.; LEISERSON, C.; RIVEST, R.; STEIN, C. **Introduction to Algorithms**. Boston: MIT Press, 2009.

DAHLBERG, I. Knowledge Organization: A New Science? **Knowl. Org.**, v. 33, n. 1, p. 11-19, 2006.

DAŠIĆ, V.; JEČMENICA, R.; LABOVIĆ, D. Functional and Information Modeling of Production Using IDEF Methods. **Strojniški vestnik - Journal of Mechanical Engineering**, v. 55, n. 2, p. 131-140, 2009.

- DAVENPORT, T. **Big Data at Work: Dispelling the Myths, Uncovering the Opportunities**. Boston: Harvard Business Review Press, 2014.
- DAVENPORT, T.; PRUSAK, L. **Conhecimento Empresarial: como as organizações gerenciam o seu capital intelectual**. Rio de Janeiro: Elsevier Science, 2003.
- DE MAURO, A.; GRECO, M.; GRIMALDI, M. A formal definition of Big Data based on its essential features. **Library Review**, v. 65, n. 3, p. 122–135, jan. 2016.
- DERVIN, B. From the mind's eye of the user: the sense-making qualitative-quantitative methodology. *In: Qualitative Research in Information Management*. Englewood: Libraries Unlimited, 1992. p. 61–84.
- DEY, A. **Providing Architectural Support for Building Context-Aware Applications**. Tese. Georgia Institute of Technology, Atlanta, nov. 2000.
- DEY, A. Understanding and Using Context. **Personal and Ubiquitous Computing**, v. 5, n. 1, p. 4–7, 2001.
- DUMONT, L. M. M.; GATTONI, R. L. C. As relações informacionais na sociedade reflexiva de Giddens. **Ciência da Informação**, v. 32, n. 3, p. 46–53, dez. 2003.
- EKBIA, H. *et al.* Big Data, Bigger Dilemmas: A Critical Review. **Journal of the Association for Information Science and Technology**, v. 66, n. 1, p. 1–34, 2015.
- FERNEDA, E. **Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. São Paulo: USP - Universidade de São Paulo, 2003.
- FERNEDA, E. **Introdução aos Modelos Computacionais de Recuperação de Informação**. Rio de Janeiro: Ciência Moderna, 2012.
- FIGUEIREDO, N. M. **Estudos de Uso e Usuários da Informação**. Brasília: IBICT, 1994.
- FORESTI, F.; VARVAKIS, G.; GODOY VIERA, A. F. A importância do contexto na Ciência da Informação. **Biblios: Journal of Librarianship and Information Science**, n. 72, p. 1–21, jul. 2016.
- FOSKETT, D. J. A note on the concept of “relevance”. **Information Storage Retrieval**, v. 8, n. 1, p. 77–78, 1972.
- FROHMAN, B. O caráter social, material e público da informação. *In: A dimensão epistemológica da informação e suas interfaces técnicas, políticas e institucionais nos processos de produção, acesso e disseminação da informação*. São Paulo: Fundepe, 2008, p.19-34.
- FRY, L.; SMITH, D. Congruence, Contingency, and Theory Building. **Academy of Management Review**, v. 12, n. 1, p. 117–132, 1987.

FUJINO, A. Acessibilidade informacional de PCD no contexto da lei de acesso à informação: desafios para estudo de usuários. **Informação em Pauta**, v. 2, p. 237-257, 2017.

GABRIEL JÚNIOR, R. F. **Socialização dos Saberes**: metodologia para desenvolvimento de um repertório temático em Ciência Da Informação. Dissertação - Universidade Federal do Paraná, Curitiba, 2011.

GÉRON, A. **Hands-on Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent System**. Nova Iorque: O'Reilly Media, 2019.

GOFFMAN, E. **Quadros da experiência social**: uma perspectiva de análise. Petrópolis: Editora Vozes, 2012.

GONZALEZ DE GÓMEZ, M. N. Regime de informação: construção de um conceito. **Inf. & Soc.:Est.**, v. 22, n. 3, p. 43-60, 2012.

GONZALEZ DE GÓMEZ, M. N.; RABELLO, R. Sujeito, agência e informação: tradição e leituras transversais. *In*: **Informação**: agentes e intermediação. Brasília: IBICT, 2017, p. 43–97.

GUINCHAT, C.; MENO, M. **Introdução geral às ciências e técnicas da informação e documentação**. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 1994.

HABERMAS, J. **Consciência moral e agir comunicativo**. Rio de Janeiro: Tempo Brasileiro, 1989.

HARRISON, M. **Machine Learning**: Guia de Referência Rápida. Nova Iorque: O'Reilly Media, 2019.

HARTER, S. P. Psychological relevance, and information science. **Journal of the American Society for Information Science**, v. 43, p. 602–615, 1992.

HJØRLAND, B. The Foundation of the Concept of Relevance. **Journal of the American Society for Information Science**, v. 61, n. 2, p. 217–237, 2010.

HOWARD, J.; GUGGER, S. **Deep Learning for Coders with Fast AI and Pytorch**: AI Applications Without a PhD. Nova Iorque: O'Reilly Media, 2020.

INGWERSEN, P. Cognitive perspectives of information retrieval interaction. **Journal of Documentation**, v. 52, n. 1, p. 3–50, 1996.

INGWERSEN, P. Cognitive information retrieval. **Annual Review of Information Science and Technology**, v. 34, p. 3–52, 1999.

INGWERSEN, P.; JÄVERLIN, K. Information seeking research needs extension towards tasks and technology. **Information Research**, v. 10, n. 1, p. 1–16, 2004.

INGWERSEN, P.; JÄVERLIN, K. **The Turn: Integration of Information Seeking and Retrieval in Context**. Berlin: Springer, 2005.

JARRAHI, M. H. et al. Artificial intelligence in the work context. **Journal of the Association for Information Science and Technology**, v. 74, p. 303–310, 2023.

JOHNSON, J. D. On contexts of information seeking. **Information Processing and Management**, v. 39, n. 1, p. 735–760, 2003.

KUHLTHAU, C. **Information search process**. Boca Raton: Encyclopedia of Library and Information Sciences, 1991.

LANCASTER, F. W. **Indexing and abstracting in theory and practice**. Londres: Library Association, 1998.

LE COADIC, Y.-F. **A ciência da informação**. Brasília: Briquet de Lemos Livros, 2004.

LIMA, G. A. DE; CAMPOS, M. L. A. Sistema de Armazenamento e Recuperação da Informação: uma análise do impacto das variáveis e medidas visando à organização e recuperação de informação centrado no usuário. **Rev. Digit. Bibl. Cienc. Inf.**, v. 20, p. 1-23, 2022.

LOCASCIO, N.; BUDUMA, N. **Fundamentals of Deep Learning: Designing Next-Generation Machine Intelligence Algorithms**. Nova Iorque: O'Reilly Media, 2017.

MANNING, C.; RAGHAVAN, P.; SCHÜTZE, H. **Introduction to Information Retrieval**. Cambridge: Cambridge University Press, 2008.

MARQUESONE, R. **Big Data: Técnicas e tecnologias para extração de valor dos dados**. São Paulo: Casa do Código, 2016.

MARTELETO, R.; NÓBREGA, N.; MORADO, D. **Cultura informacional: demarcações de uma linha de estudos de cultura, informação e sociedade**. Em: fronteiras da ciência da informação. Brasília: IBICT, 2013, p. 78-106.

MINSKY, M. **The Society of Mind**. Nova Iorque: Simon & Schuster, 1986.

MIZZARO, S. How many relevances in information retrieval? **Interacting with Computers**, v. 10, p. 303–320, 1998.

MORVILLE, P.; ROSENFELD, L. **Information Architecture for the World Wide Web**. Nova Iorque: O'Reilly Media, 2006.

NASSIF, M. E.; VENÂNCIO, L. S.; HENRIQUE, L. C. J. Sujeito, contexto e tarefa na busca de informação: uma análise sob a ótica da cognição situada. **DataGramZero - Revista de Ciência da Informação**, v. 8, n. 8, p. 1–12, out. 2007.

NEELAKANTAN, A. et al. Text and Code Embeddings by Contrastive Pre-Training. **arXivLabs**, v. 1, n. 1, p. 1-13, 2022.

NEVELING, U.; WERSIG, G. The phenomena of interest to Information Science. **Information Scientist**, v. 9, n. 4, p. 127–140, dez. 1975.

NORMAN, D. **O Desgin do dia a dia**. Rio de Janeiro: Rocco, 1986.

NORVIG, P.; RUSSELL, S. **Inteligência Artificial**. Rio de Janeiro: Campus, 2004.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37 – 44, 1996.

O'NEIL, C. **Algoritmos de Destruição em Massa**. Santo André: Rua do Sabão, 2021.

OLIVEIRA, G. L. C. DE; RAMALHO, F. A. Necessidades e uso de informação na prática docente. *In*: ENANCIB 2012. **Anais [...]**. Rio de Janeiro, 2012, p. 1-19.

PONJUÁN-DANTE, G. **Gestión de información dimensiones e implementación para el éxito organizacional**. Buenos Aires: Nuevo Paradigma Ediciones, 2006.

RABELLO, R. Noções de sujeito em modelos teóricos na ciência da informação: do enfoque no sistema à consideração da agência em contexto. **Informação e Sociedade**, v. 23, n. 3, p. 57–71, 2013a.

RABELLO, R. Leituras sobre usuário e uso de informação na Ciência da Informação. **Perspectivas em Ciência da Informação**, v. 18, n. 4, p. 152–184, dez. 2013b.

RABELLO, R. Sujeito e agência informacional: comportamento, prática e ação. *In*: **Informação: agentes e intermediação**. Brasília: IBICT, 2017, p. 101-152.

RABELLO, R.; GONZALEZ DE GÓMEZ, M. N. Agentes, intermediações e institucionalidades: apontamentos acerca de um mosaico interpretativo no campo informacional. *In*: **Informação: agentes e intermediação**. Brasília: IBICT, 2017, p. 21-42.

RECKER, J. et al. From Representation to Mediation: A New Agenda for Conceptual Modeling Research In A Digital World. **MIS Quarterly**, v. 45, n. 1, p. 269–300, 2021.

REID, J. A New, task-oriented paradigm for information retrieval: Implications for evaluation of information retrieval systems. *In*: **Proceedings of CoLIS 3, third international conference on the conceptions of library and information science**. Lovke: Naklada Benja, 1999. p. 97–108.

ROWLEY, J. **Abstracting and indexing**. Londres: Clive Bingley, 2002.

ROWLEY, J. The wisdom hierarchy: representations of the DIKW hierarchy. **Journal of Information Science**, v. 33, n. 2, p. 163–180, abr. 2007.

SANZ CASADO, E. **Manual de estudos de usuarios**. Madri: Fundación Germán Sánchez Ruipérez, 1994.

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**, p. 321–343, 1975.

SARACEVIC, T. Modeling interaction in information retrieval: a review and proposal. **Proceedings of the Annual Academy Meeting of American Society for Information Science**, v. 33, p. 3–9, 1996.

SAVOLAINEN, R. Information behavior and information practice: reviewing the 'umbrella concepts' of information-seeking studies. **Library Quarterly**, v. 77, n. 2, p. 109 – 132, 2007.

SAVOLAINEN, R.; THOMSON, L. Assessing the theoretical potential of an expanded model for everyday information practices. **Journal of the Association for Information Science and Technology**, v. 73, n. 4, p. 511–527, 2022.

SCHAMBER, L.; EISENBERG, M.; NILAN, M. A re-examination of relevance: toward a dynamic, situational definition. **Information Processing and Management**, v. 26, n. 6, p. 755–776, 1990.

SEVERINO, A. **Metodologia do Trabalho Científico**. São Paulo: Cortez, 2013.

SILVA, L. R. A. B. Inteligência artificial em processos de extração de conhecimento KDD e KDT. **REU - Revista de Estudos Universitários**, v. 46, n. 1, p. 161–180, 2020.

SMIT, J. W. A informação na Ciência da Informação. **InCID: Revista de Ciência da Informação e Documentação**, v. 3, n. 2, p. 84–101, dez. 2012.

SMIT, J. W.; BARRETO, A. DE A. Ciência da informação: base conceitual para a formação do profissional. *In: Formação do profissional da informação*. São Paulo: Polis, 2002. p. 9–23.

SOUZA, R. R.; ALMEIDA, M. B.; BARACHO, R. M. A. Ciência da informação em transformação: Big Data, nuvens, redes sociais e Web Semântica. **Big Data**, v. 42, n. 2, p. 159 - 173, 2013.

TALJA, S.; HEIDI, K.; TARJA, P. The production of 'context' in information seeking research: a metatheoretical view. **Information Processing and Management**, v. 35, n. 1, p. 751–763, 1999.

TARAPANOFF, K. Informação, conhecimento e inteligência em corporações: relações e complementaridade. *In: Inteligência, Informação e Conhecimento*. Brasília: Instituto Brasileiro de Informação em Ciência e Tecnologia, 2006. p. 19–35.

TOCCI, R.; WIDMER, N.; MOSS, G. **Digital Systems**. Nova Iorque: Pearson, 2016.

VASWANI, A. et al. Attention Is All You Need. **31st Conference on Neural Information Processing Systems**, p. 1–15, 2017.

VICKERY, A.; VICKERY, B. **Information Science in Theory and Practice**. Munique: K.G. Saur, 2004.

VIDAL, A. **Análise Preditiva de Dados/FEA-USP/EAD5754-10/3**. São Paulo, 2019.

WEBER, M. Ação social e relação social. *In: Sociologia e sociedade: leituras de introdução à sociologia*. Rio de Janeiro: LTC, 1977, p. 126-134.

WERSIG, G.; WINDEL, G. Information Science needs a theory of information actions. **Social Science Information Studies**, v. 5, n. 1, p. 11–23, 1985.

WHITE, T. **Hadoop, the definitive guide: Storage and Analysis at Internet Scale**. Sebastopol: O'Reilly Media, 2015.

WILSON, T. Human information behavior. **Informing Science**, v. 3, n. 2, p. 48 – 55, 2000.

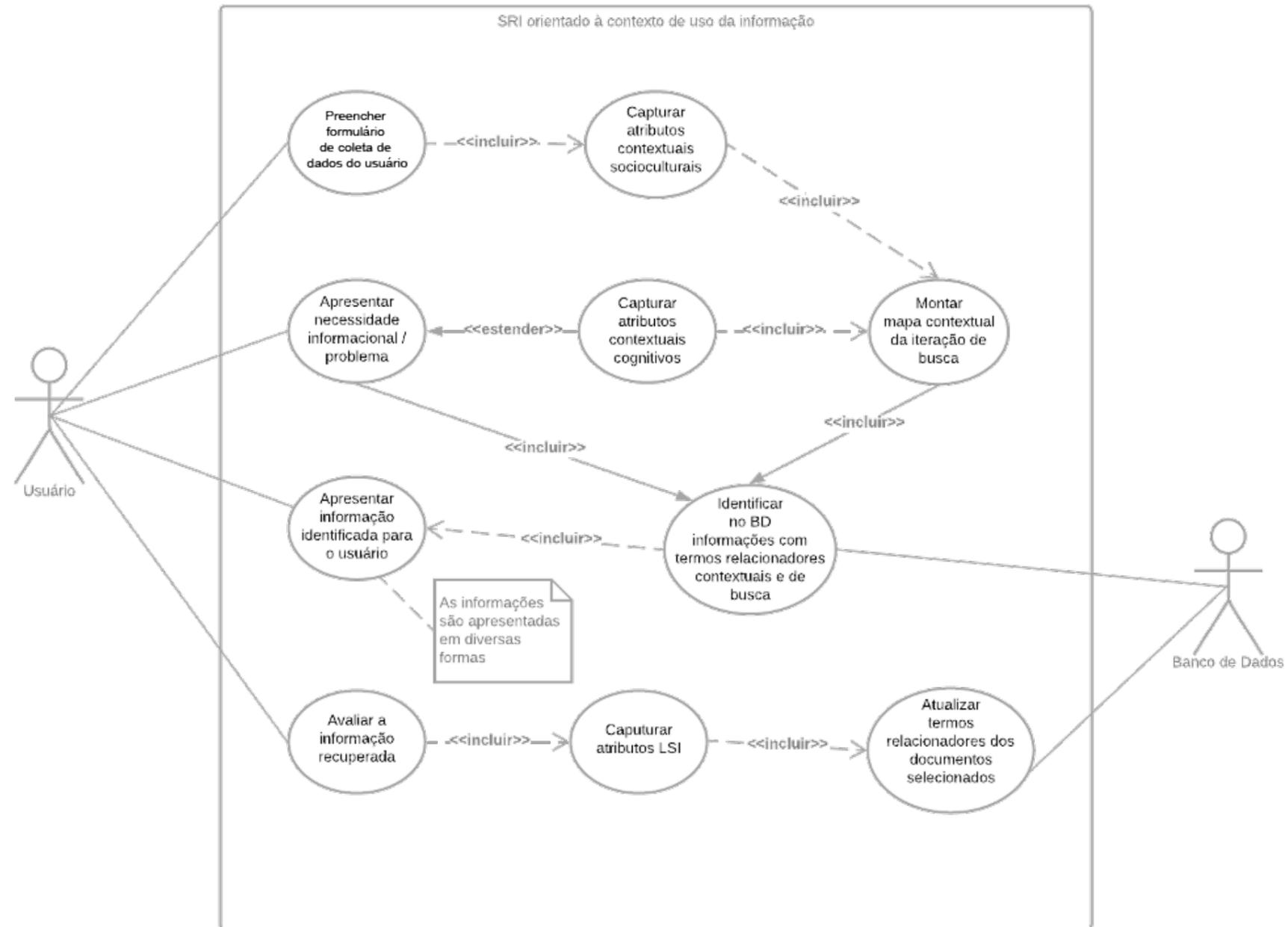
WILSON, T. On User Studies and Information Needs. **Journal of Documentation**, v. 62, n. 6, p. 658-670, 2006.

ZANT, T. VAN DER; SCHOMAKER, L.; KOUW, M. Generative Artificial Intelligence. *In: Theory and Philosophy of Artificial Intelligence*. Nova Iorque: Springer, 2012.

ZUBOFF, S. **A Era do Capitalismo de Vigilância**. São Paulo: Intrínseca, 2021.

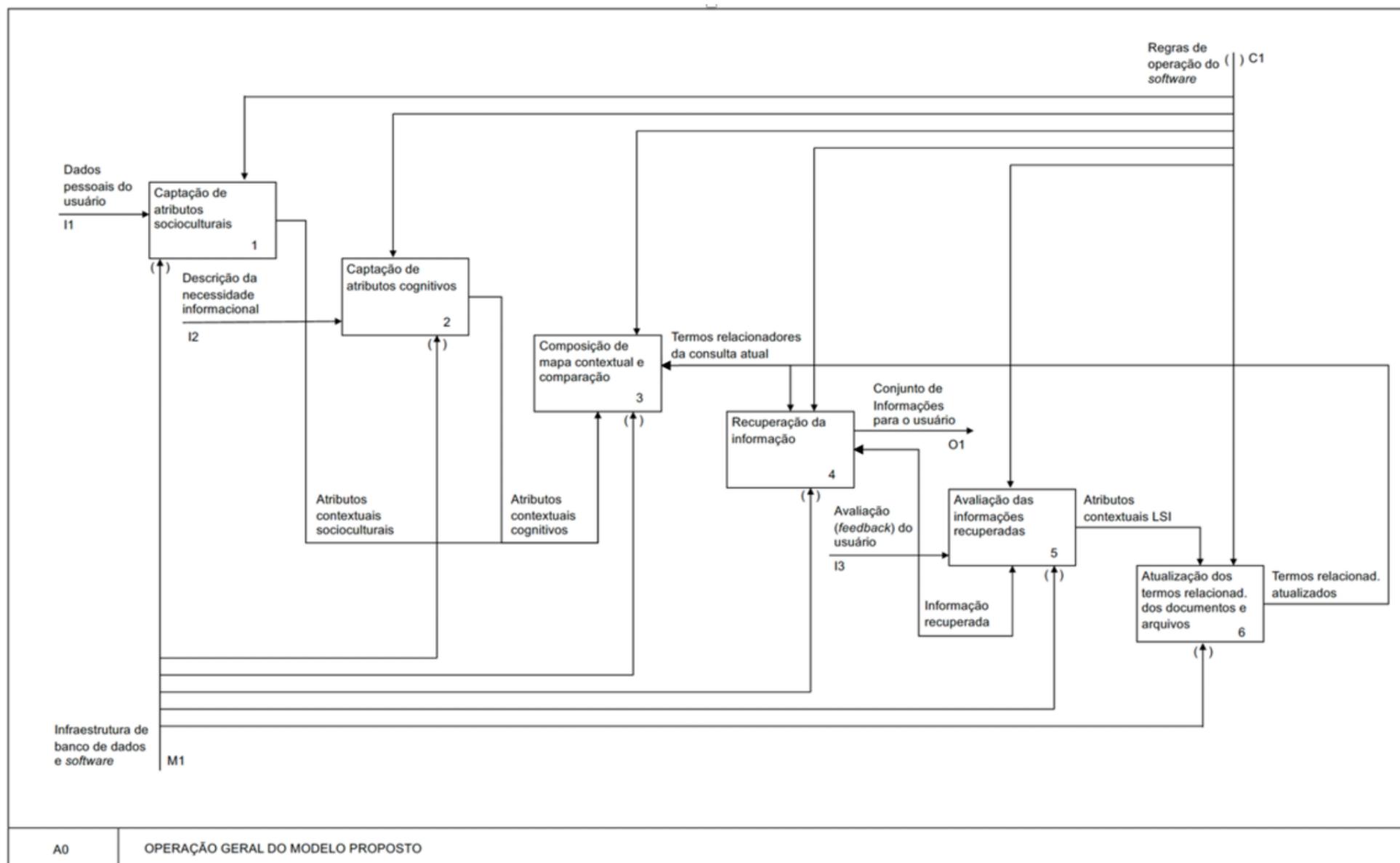
## **APÊNDICES**

Apêndice 1



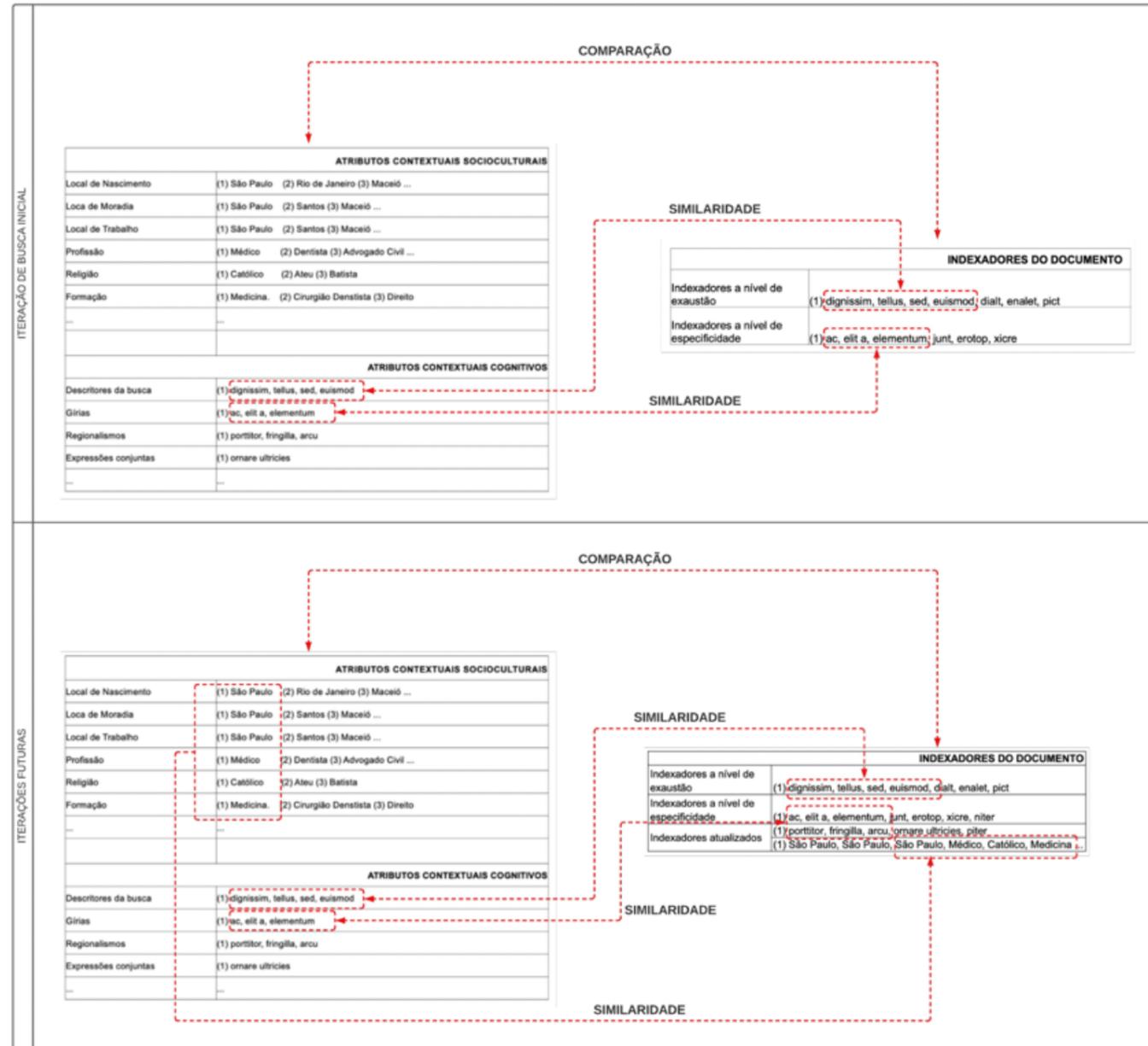
Fonte: Elaborado pelo autor.

Apêndice 2



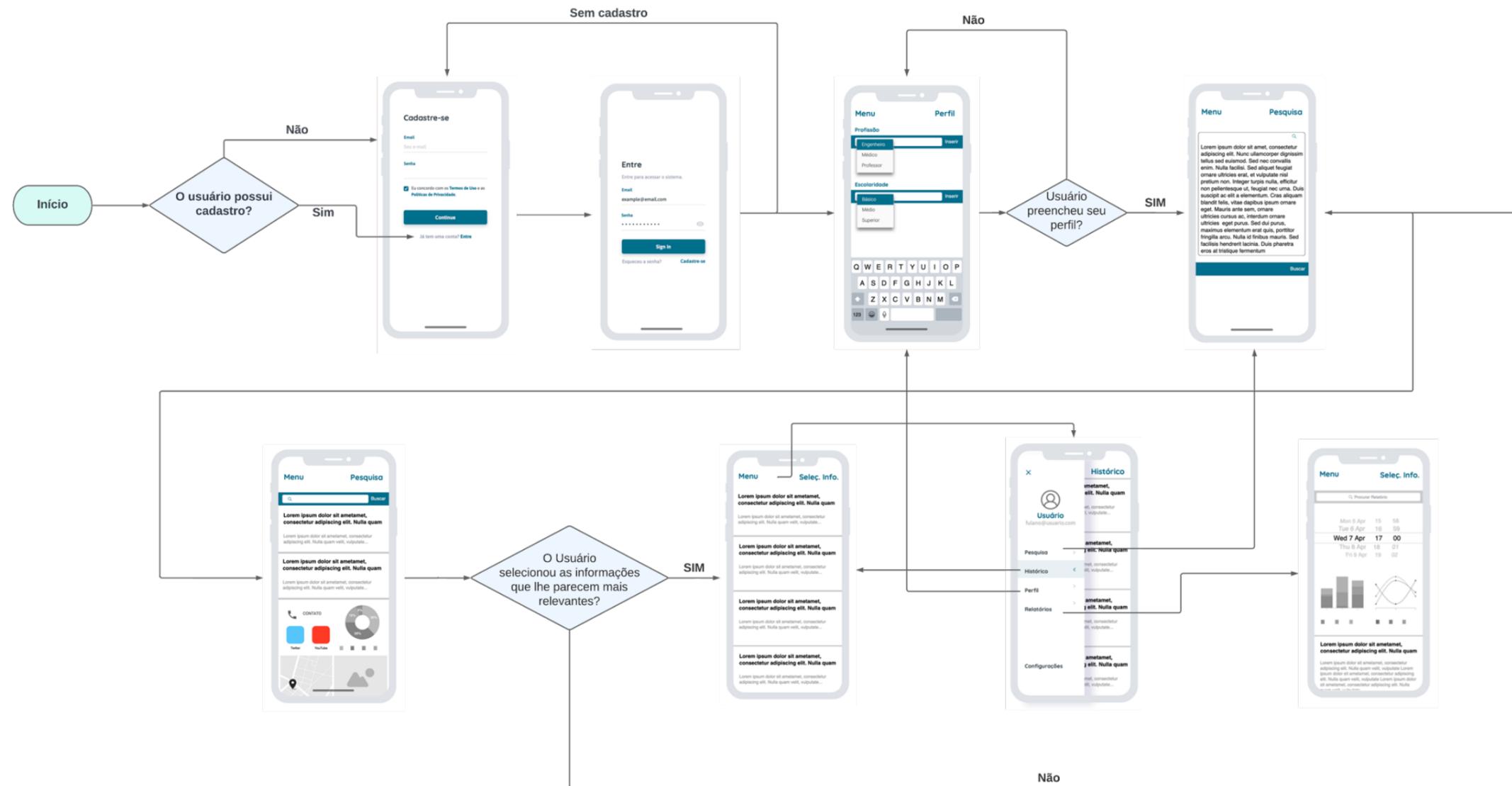
Fonte: Elaborado pelo autor.

Apêndice 3



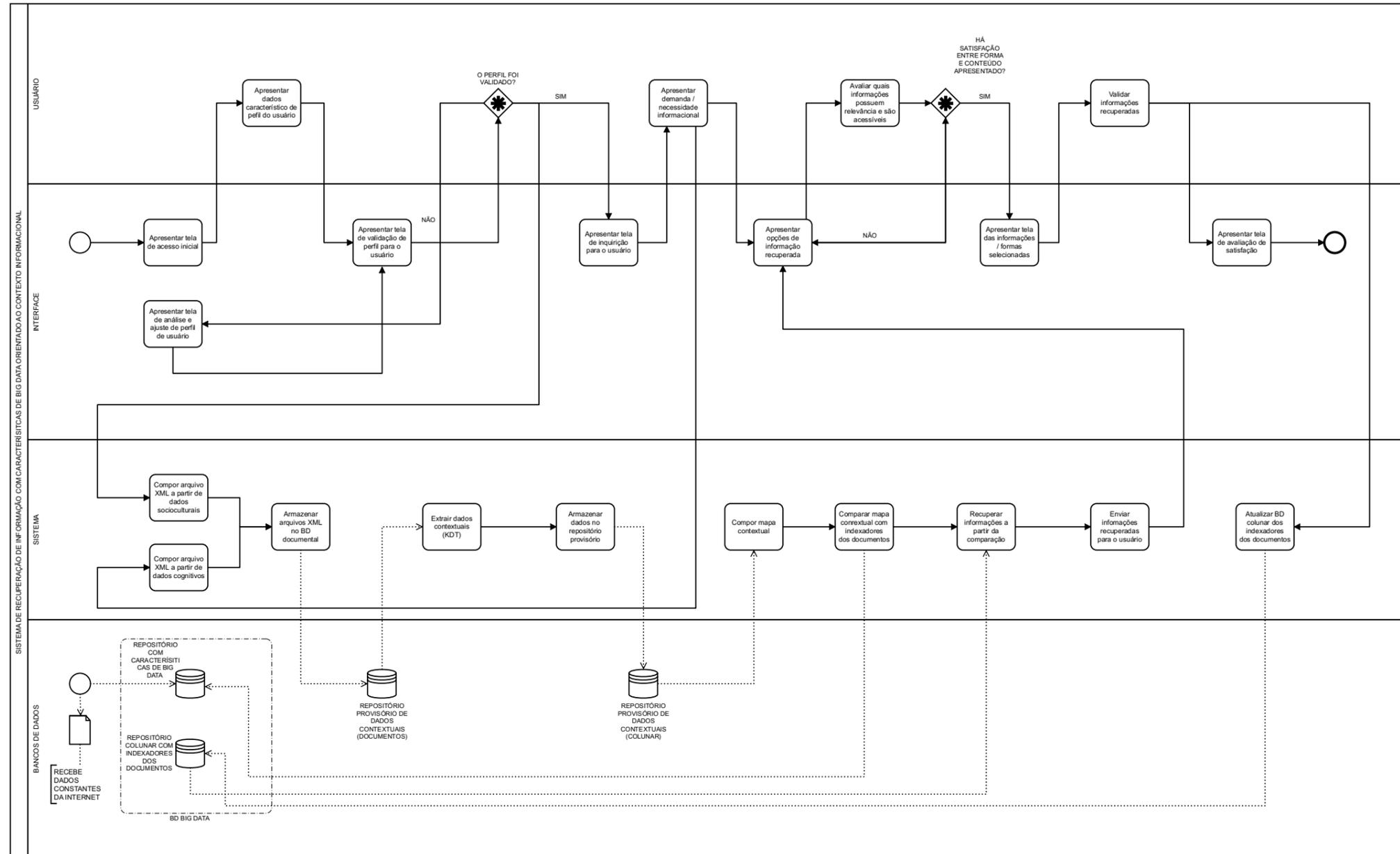
Fonte: Elaborado pelo autor.

# Apêndice 4



Fonte: Elaborado pelo autor.

## Apêndice 5



Fonte: Elaborado pelo autor.