

UNIVERSIDADE ESTADUAL PAULISTA
FACULDADE DE FILOSOFIA E CIÊNCIAS
CAMPUS MARÍLIA
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA INFORMAÇÃO

FABRICIO BAPTISTA

**UMA PROPOSTA DE INTERFACE DE RESULTADOS DE
BUSCAS EM SISTEMAS DE RECUPERAÇÃO DE
INFORMAÇÃO: A SEMIÓTICA E A INTERAÇÃO HUMANO
COMPUTADOR COMO APORTE TEÓRICO**

Marília
2019

FABRICIO BAPTISTA

**UMA PROPOSTA DE INTERFACE DE RESULTADOS DE
BUSCAS EM SISTEMAS DE RECUPERAÇÃO DE
INFORMAÇÃO: A SEMIÓTICA E A INTERAÇÃO HUMANO
COMPUTADOR COMO APORTE TEÓRICO**

Tese apresentada ao Programa de Pós-Graduação em
Ciência da Informação da Universidade Estadual Paulista
(UNESP), Faculdade de Filosofia e Ciências de Marília,
como requisito parcial para obtenção do título de Doutor
em Ciência da Informação.

Linha de Pesquisa: Informação e Tecnologia.

Orientador: Prof. Dr. Edberto Ferneda.

Coorientador: Prof. Dr. Carlos Cândido de Almeida.

Marília
2019

B222p

Baptista, Fabricio

Uma proposta de interface de resultados de buscas em sistemas de recuperação de informação : a semiótica e a interação humano computador como aporte teórico / Fabricio Baptista. -- Marília, 2019
171 p.

Tese (doutorado) - Universidade Estadual Paulista (Unesp),
Faculdade de Filosofia e Ciências, Marília

Orientador: Edberto Ferneda

Coorientador: Carlos Cândido de Almeida

1. Recuperação de Informação. 2. Interação Homem-Maquina. 3.
Interfaces (Computador). 4. Semiótica. I. Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca da Faculdade de Filosofia e Ciências, Marília. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

FABRÍCIO BAPTISTA

**UMA PROPOSTA DE INTERFACE DE RESULTADOS DE
BUSCAS EM SISTEMAS DE RECUPERAÇÃO DE
INFORMAÇÃO: A SEMIÓTICA E A INTERAÇÃO HUMANO
COMPUTADOR COMO APORTE TEÓRICO**

BANCA EXAMINADORA

Prof. Dr. Edberto Ferneda (Orientador)
Departamento de Ciência da Informação
Universidade Estadual Paulista (UNESP) – Campus de Marília

Prof. Dr. Walter Moreira
Departamento de Ciência da Informação
Universidade Estadual Paulista (UNESP) – Campus de Marília

Prof. Dra. Maria José Vicentini Jorente
Universidade Estadual Paulista (UNESP) – Campus de Marília

Prof. Dra. Ana Carolina Simionato
Universidade Federal de São Carlos (UFSCar)

Prof. Dr. José Remo Ferreira Brega
Universidade Estadual Paulista - Faculdade de Ciências - Campus de Bauru

Marília, 27 de fevereiro de 2019.

Dedico este trabalho à minha Família!

AGRADECIMENTOS

A Deus por conduzir os meus passos trilhando caminhos que muitas vezes não entendi, mas que com certeza me levaram a ser uma pessoa melhor.

A minha família, em especial aos meus filhos Fabrício, Isabella e Arthur e minha esposa Daniele pela compreensão nos momentos em que estive ausente, e também pelo carinho nos momentos mais difíceis.

Aos meus pais, Francisco e Ione, que mesmo distantes sempre acreditaram em mim e me incentivaram a lutar pelos meus sonhos.

Aos meus irmãos Rodrigo e Álvaro (*in memoriam*) pelo carinho durante os anos de convivência.

Ao Prof. Dr. Edberto Ferneda, meu orientador e que sempre me apoiou em todas as horas, agradeço pelas orientações e conselhos durante todo o período de estudo.

Ao Prof. Dr. Carlos Cândido de Almeida por também me orientar e apresentar uma área profícua de estudo que rendeu ótimos frutos durante minhas pesquisas.

Aos professores que contribuíram para minha formação no programa de doutorado, em especial aos professores Dr. Walter Moreira e Dra. Maria José Vicentini Jorente, os quais me proporcionaram realizar estudos muito proveitosos durante suas disciplinas e aceitaram participar da minha defesa.

Aos professores Dr. José Remo Ferreira Brega e Dra. Ana Carolina Simionato, membros da banca avaliadora, por aceitarem fazer parte da banca de defesa.

A todos meus colegas de doutorado que me acompanharam durante as disciplinas e proporcionaram uma troca de experiências fantástica.

Ao Instituto Federal do Paraná – IFPR pelo incentivo institucional à minha formação e ao meu afastamento.

À Fundação Araucária pelo apoio financeiro que proporcionou aprofundar minha pesquisa através da aquisição de livros e participação em eventos.

À todos que participaram direta ou indiretamente da conclusão desta etapa gratificante de minha trajetória.

Uma máquina pode fazer o trabalho de cinquenta pessoas comuns. Máquina alguma pode fazer o trabalho de um homem incomum. (Elbert Hubbard)

O único homem que não erra é aquele que nunca fez nada. (Roosevelt)

BAPTISTA, Fabrício. **Uma proposta de interface de resultados de buscas em sistemas de recuperação de informação: a semiótica e a interação humano computador como aporte teórico**. 171. Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação – Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2019.

RESUMO

Com o crescente aumento de documentos digitais disponíveis em acervos on-line, torna-se inevitável realizar estudos que proporcionem avanços no modo como os usuários buscam informações nesses documentos. A Recuperação de Informação procura meios para aperfeiçoar este processo através de diferentes núcleos de estudo como modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuário, visualização de dados, entre outros. Diante dessas opções passíveis de investigação, optou-se por restringir esta pesquisa às interfaces de usuário, especificamente os resultados de busca. Nesta perspectiva, o objetivo deste estudo é propor um modelo de interface computacional para apresentação dos resultados de busca em sistemas de recuperação de informação, agregando elementos visuais fundamentados nas áreas de Semiótica e da Interação Humano Computador. Neste sentido, realizou-se uma pesquisa dividida em duas partes. Em um primeiro momento, apresenta-se uma pesquisa bibliográfica de caráter exploratório. Em seguida, foi realizada a pesquisa aplicada. A pesquisa bibliográfica embasou a fundamentação teórica e forneceu subsídios para o levantamento de desafios da recuperação de informação. A pesquisa aplicada procurou responder de forma prática ao objetivo e questionamento da pesquisa. Assim, apresenta-se uma proposta de interface que combina recursos visuais e textuais para contribuir com o processo de percepção dos usuários em relação aos resultados das buscas. Concluindo, acredita-se que é possível relacionar ao desenvolvimento de interfaces de Recuperação de Informação os recursos visuais e de interação fundamentados na semiótica e viabilizados pela Interação Humano Computador.

Palavras-chave: Recuperação de Informação. Interação Humano Computador. Semiótica. Interfaces.

BAPTISTA, Fabrício. **A proposal of interface search results in information retrieval systems: the semiotics and human computer interaction as a theoretical contribution.** 171. Tese apresentada ao Programa de Pós-Graduação em Ciência da Informação – Faculdade de Filosofia e Ciências da Universidade Estadual Paulista – UNESP, Marília, 2019.

ABSTRACT

With the growing number of digital documents available in online collections, it becomes inevitable to conduct studies that provide advances in how users search for information in those documents. Information Retrieval looks for ways to improve this process through different study centers such as modeling, text classification, system architecture, user interfaces, data visualization, among others. Faced with these investigative options, it was decided to restrict this research to user interfaces, specifically search results. In this perspective, the objective of this study is to propose a computational interface model to present the search results in information retrieval systems, adding visual elements based on the Semiotics and Human Computer Interaction areas. In this sense, a research was carried out in two parts. At first, an exploratory bibliographic research is presented. The applied research was then carried out. The bibliographical research based the theoretical foundation and provided subsidies for the collection of information retrieval challenges. The applied research sought to respond in a practical way to the objective and questioning of the research. Thus, a proposal of interface is presented that combines visual and textual resources to contribute to the process of perception of the users in relation to the search results. In conclusion, it is believed that it is possible to relate to the development of Information Retrieval interfaces the visual and interaction resources based on semiotics and made possible by Human Computer Interaction.

Palavras-chave: Information Retrieval. Human Computer Interaction. Semiotics. Interfaces.

LISTA DE FIGURAS

Figura 1: Representação do processo de recuperação de informação	24
Figura 2: Modelo de comportamento informacional de Wilson	33
Figura 3: Estrutura do modelo do <i>sense-making</i> de Dervin	33
Figura 4: Metáfora do modelo <i>sense-making</i> de Dervin	34
Figura 5: Fases do comportamento na busca informacional de Ellis	35
Figura 6: Número de termos utilizados nas pesquisas do Excite	40
Figura 7: Número de termos utilizados nas consultas relatadas pelos pós-graduandos	41
Figura 8: Discrepância entre o mundo psicológico e o físico.....	57
Figura 9: Golfos de execução e avaliação.	58
Figura 10: Interface baseada em clusters.....	62
Figura 11: Interface do <i>SoundAnchoring</i>	65
Figura 12: Exemplos das técnicas de visualização exploradas SPIRE, gCLUTO e Infosky ...	70
Figura 13: Exemplos de técnicas de visualização	70
Figura 14: Exemplo de visualização radial.....	72
Figura 15: Projeto filosófico peirceano	75
Figura 16: Constituintes peirceanos da percepção	84
Figura 17: Elementos do diagrama de Casos de Uso	94
Figura 18: Diagrama de caso de uso da proposta	95
Figura 19: Exemplo e notação gráfica do diagrama de classes	97
Figura 20: Diagrama de classes da proposta	98
Figura 21: Exemplo de documentos indexados.....	99
Figura 22: Arquitetura do software	101
Figura 23: Componentes Solr - Indexador e Buscador.....	103
Figura 24: Representação vetorial de um documento com três termos de indexação.....	105
Figura 25: Exemplo de indexação no Solr.....	106
Figura 26: Representação vetorial de uma expressão de busca.....	107
Figura 27: Exemplo de busca no Solr.....	108
Figura 28: Exemplos de resposta do Solr	109
Figura 29: Arquitetura cliente-servidor	110
Figura 30: Arquitetura cliente-servidor da proposta.....	110
Figura 31: Exemplo HTML.....	112
Figura 32: Exemplo de estilo utilizando CSS.....	113

Figura 33: Exemplo de função JavaScript.....	114
Figura 34: Exemplo de instruções em PHP	116
Figura 35: Tela inicial de Busca.....	117
Figura 36: Código HTML e CSS da página de Busca.....	118
Figura 37: Área de resultados de busca	119
Figura 38: Código PHP	120
Figura 39: Matriz de resultados do Solr	120
Figura 40: Detalhes do documento.....	121
Figura 41: Acesso ao documento completo.....	122
Figura 42: Pesquisa a partir de palavra-chave	122
Figura 43: Interface utilizando a técnica Radviz	125
Figura 44: Detalhe das palavras-chave.....	126
Figura 45: Estilos de interação presentes no protótipo.....	127
Figura 46: Utilização de ícones, índices e símbolos.....	129
Figura 47: Dimensões das inferências abduativas.....	130
Figura 48: Exemplo de escala de diferencial semântico.....	136
Figura 49: Tabela de execução das tarefas	141
Figura 50: Proposta de alteração de interface.....	142
Figura 51: Mapa de calor da interação dos usuários na busca.....	145
Figura 52: Mapa de calor da interação dos usuários com os documentos recuperados	145

LISTA DE QUADROS

Quadro 1: Percurso metodológico da pesquisa	19
Quadro 2: Aspectos de relevância	45
Quadro 3: Trabalhos que abordaram problemas na RI.....	49
Quadro 4: Ferramentas de visualização.....	69
Quadro 5: Teorias sobre sensação e percepção	82
Quadro 6: Casos de uso do ator usuário	96
Quadro 7: Relação entre os diagramas de casos de uso e classes.....	100
Quadro 8: Percepções esperadas para cada caso de uso	133
Quadro 9: Cenário para o roteiro de teste.....	135
Quadro 10: Exemplo de perguntas do questionário de avaliação.....	137
Quadro 11: Recursos para o plano de testes	138
Quadro 12: Síntese do perfil dos participantes	140
Quadro 13: Relação de afirmativas em relação às tarefas.	144
Quadro 14: Escala semântica de usabilidade.....	147

LISTA DE ABREVIATURAS E SIGLAS

CAPES	Coordenação de Aperfeiçoamento de Pessoal de Nível Superior
CI	Ciência da Informação
CSS	<i>Cascading Style Sheets</i>
ES	Engenharia Semiótica
GOMS	<i>Goals, Operators, Methods and Selection rules</i>
HTML	<i>Hyper Text Mark Up Language</i>
HTTP	<i>HyperText Transfer Protocol</i>
IDF	<i>Inverse Document Frequency</i>
IHC	Interação Humano Computador
IN-SPIRE	<i>Spatial Paradigm FOR Information Retrieval</i>
ISP	<i>Information Search Process</i>
JSON	<i>JavaScript Object Notation</i>
PDF	<i>Portable Document Format</i>
PHP	<i>Hypertext Preprocessor</i>
RI	Recuperação de Informação
SOM	<i>Self-Organization Maps</i>
SRI	Sistemas de Recuperação de Informação
TF	<i>Term Frequency</i>
UML	<i>Unified Modeling Language</i>
VI	Visualização de Informação
WIMP	<i>Windows, Icons, Menus, and Pointers</i>
XML	<i>Extensible Markup Language</i>

SUMÁRIO

1 INTRODUÇÃO.....	14
1.1 Delimitação do Tema e Justificativa.....	15
1.2 Problema da Pesquisa	17
1.3 Hipóteses.....	17
1.4 Tese	17
1.5 Objetivos.....	17
1.6 Metodologia.....	18
1.7 Estrutura da tese.....	20
2 RECUPERAÇÃO DE INFORMAÇÃO	22
2.1 Aspectos históricos	22
2.2 O processo de Recuperação de Informação	23
2.2.1 Representação de documentos	24
2.2.2 Usuários na Recuperação de Informação	29
2.2.3 Modelos de Recuperação de Informação.....	41
2.2.4 Aspectos de Relevância na Recuperação de Informação	45
2.3 Desafios da Recuperação de Informação.....	47
2.4 Sistematização e discussão	49
3 INTERAÇÃO HUMANO COMPUTADOR - IHC.....	51
3.1 Fatos históricos	51
3.2 Conceitos básicos de IHC	52
3.2.1 Interfaces.....	52
3.2.2 Interação	53
3.3 Bases Teóricas da IHC.....	56
3.3.1 A Engenharia Cognitiva	56
3.3.2 A Engenharia Semiótica	58
3.4 Produções científicas sobre Interfaces de Recuperação de Informação	60
3.5 Visualização de Informação	66
3.6 Sistematização e discussão	73
4 SEMIÓTICA PEIRCEANA PARA O DESENVOLVIMENTO DE INTERFACES ...	74
4.1 Semiótica Peirceana.....	74
4.1.1 Fenomenologia	75

4.1.2 Peirce e o Raciocínio Diagramático	78
4.2 Teoria da percepção em Peirce e as contribuições para o desenvolvimento de interfaces	80
4.2.1 O processo de percepção em Peirce.....	83
4.2.2 A percepção e o desenvolvimento de interfaces de Recuperação de Informação	87
4.3 Sistematização e Discussão	90
5 PROPOSTA DE INTERFACE DE RESULTADO DE BUSCA.....	92
5.1 Modelagem do protótipo utilizando a UML.....	92
5.1.1 Diagrama de casos de uso.....	93
5.1.2 Diagrama de classes.....	96
5.2 Tecnologias utilizadas na implementação do protótipo.....	101
5.2.1 Motor de busca Solr.....	102
5.2.2 Tecnologias para interfaces Web.....	109
5.3 Protótipo	116
5.4 Discussão da interface de resultados de busca	123
5.4.1 Disposição dos elementos através da técnica Radviz e o modelo vetorial	124
5.4.2 Estilos de interação do protótipo	126
5.4.3 Aporte da semiótica para a concepção do protótipo.....	128
5.5 Usabilidade	131
5.5.1 Delimitação do teste de Usabilidade: determinar, explorar e escolher.....	133
5.5.2 Estruturação do teste: "identificar" e "decidir"	137
5.5.3 Aplicação e análise: avaliar	139
5.6 Sistematização e Discussão	147
6 CONSIDERAÇÕES FINAIS.....	149
REFERÊNCIAS	152
APÊNDICES	165
APÊNDICE A – Roteiro de teste de usabilidade	166
APÊNDICE B – Questionário de usabilidade	167
APÊNDICE C – Carta de consentimento	169
APÊNDICE D – Tabulação dos dados	170

1 INTRODUÇÃO

A crescente expansão do volume de informações *on-line*, faz com que os sistemas de busca tornem-se cada vez mais necessários, tornando mais complexa a resposta às necessidades de informação dos usuários modernos. Dentre as áreas que buscam abordar problemas de acesso à informação está a Recuperação de Informação (RI). Conforme descrito por Ferneda (2013), a RI envolve dois componentes principais, ou seja, de um lado as pessoas com necessidades informacionais que buscam por informação, e do outro lado os acervos documentais compostos por objetos informacionais.

Os processos de interação entre usuários e documentos são operacionalizados por Sistemas de Recuperação de Informação (SRIs). Vieira e Correa (2011, p. 74), consideram que os SRIs "são *softwares* que visam satisfazer as necessidades informacionais dos usuários, auxiliando-os a recuperar conteúdo pertinente no grande volume de informações disponíveis em coleções de documentos". Souza (2006, p. 162) complementa explicando que um sistema de Recuperação de Informação apenas informa sobre a existência de documentos relacionados à sua requisição, ficando a cargo do usuário filtrar os documentos realmente relevantes.

Alguns fatores importantes na recuperação de informação são elencados por Huang (1999, p. 2). O autor explica que a grande diversidade de documentos torna a informação bastante heterogênea. Também salienta que a diversidade dos usuários apresenta grande variação em suas necessidades, expectativas e conhecimento, além disso apresenta uma estimativa que cerca de 85% dos usuários que utilizam sites de busca, não olham a segunda tela de resultados.

Considerando tais aspectos, nota-se que a exibição de informações precisas e de forma objetiva é essencial em sistemas de recuperação de informação. Muito tem se discutido sobre algumas limitações dos sistemas de recuperação de informação; aspectos como relevância, semântica e sobrecarga de informação são frequentemente citados como obstáculos a serem superados.

Saracevic (1996) descreve que problemas complexos requerem enfoques interdisciplinares e soluções multidisciplinares. Para o autor, a interdisciplinaridade foi iniciada na Ciência da Informação quando se tentou resolver alguns problemas básicos sobre a compreensão da informação e a comunicação, suas manifestações, o comportamento informativo humano e os problemas aplicados ligados à acessibilidade de informação. Seguindo o panorama, o surgimento da área de Recuperação de Informação (apresentado

originalmente por Mooers (1951)) ocorreu na relação entre as Ciências da Informação e da Computação, sendo que um dos propósitos desta área é abordar problemas de acesso à informação. Diante do exposto, pressupõe-se que os SRIs podem ser combinados com outras tecnologias, com o intuito de aumentar a eficiência de suas respostas.

Nesta perspectiva interdisciplinar, serão observadas as áreas de Interação Humano Computador (IHC) e Semiótica. O processo de significação vislumbrado pela semiótica também pode ser aplicado na organização e representação do conhecimento. Para Barros e Café (2012), esses processos de significação são aspectos primordiais no desenvolvimento de métodos para interpretar diferentes constituições textuais. Andersen (2001) também relaciona o núcleo da semiótica com estudos focados na interação entre máquinas e seres humanos, argumentando que assim é possível discutir representações e também a forma como os usuários interpretam essas representações.

Os objetivos da IHC e da RI estão em consonância no que diz respeito a investigar meios que facilitem a comunicação dos usuários com sistemas automatizados. A área de IHC não se refere apenas às formas de interação entre usuários e computadores, mas principalmente às teorias e técnicas de projeto de sistemas interativos. Através dos conceitos e técnicas estruturadas pela IHC, é possível formar uma base conceitual que forneça apoio na produção de um sistema de RI em conjunto com elementos fundamentais da Semiótica..

Esse trabalho concentra-se na da reflexão sobre a utilização de elementos visuais nas interfaces de recuperação de informação. Procura ponderar sobre a aplicação das bases da IHC e Semiótica como apoio aos sistemas de recuperação de informação.

1.1 DELIMITAÇÃO DO TEMA E JUSTIFICATIVA

A recuperação de informação possui vários núcleos de estudos que atuam na melhoria dos processos de busca como modelagem, classificação de textos, arquitetura de sistemas, interfaces de usuário, visualização de dados, filtragem e linguagens (BAEZA YATES e RIBEIRO NETO, 2013, p. 1). Para Ahn e Brusilovsky (2013), é necessário a realização de estudos que visem suportar a crescente complexidade das tarefas de busca. Os pesquisadores de recuperação de informação têm buscado uma variedade de abordagens com o objetivo de proporcionar aos usuários interfaces mais poderosas para apoiar o processo de pesquisa. Diante dessas opções passíveis de investigação, optou-se por restringir esta análise às interfaces de usuário, especificamente os resultados de busca.

Singh, Hsu e Moon (2013) fomentam a possibilidade de estudos dessas

interfaces ao afirmar que o processo atual de visualização de resultados das buscas é insuficiente para descrever as informações que o usuário precisa, bem como a conteúdo que supostamente possa satisfazê-lo. Hobbs, Pfitzner e Powers (2002) apresentaram algumas desvantagens desse formato de interface:

- a) Os usuários devem inspecionar sequencialmente uma lista geralmente longa de documentos;
- b) O usuário geralmente não é informado sobre a razão pela qual o documento foi recuperado;
- c) O usuário não é informado sobre a relação de um determinado documento com a busca;
- d) Não são representadas as semelhanças entre os documentos; e
- e) É exibido apenas um pequeno trecho do documento para que o usuário possa julgar a relevância.

Para abordar esse problema, faz-se necessário agregar outras áreas, além da RI, que possam aprimorar o processo de busca informacional. A temática aqui apresentada propõe um modelo de interface de resultados de busca em um sistema de RI que adota a disposição de elementos visuais em conjunto com os elementos textuais, para representar documentos e sua relevância em relação às requisições dos usuários. No que tange à utilização dos elementos visuais, recorre-se às bases teóricas desenvolvidas pela semiótica peirceana e os fundamentos da IHC. Esses conceitos serão detalhados apropriadamente nos próximos capítulos.

O aporte da IHC é operacionalizado através dos estudos de interação e de técnicas de visualização; assim são combinados alguns estilos de interação já conhecidos pelos usuários, além de trabalhar a disposição dos elementos na interface. Já a semiótica apresenta a importância dos signos no processo de percepção das interfaces, destacando o potencial comunicativo dos elementos visuais. Em resumo, destacam-se o modelo vetorial da RI, o modelo de pensamento diagramático de Peirce e a semiótica, e os conceitos de interação e visualização destacados na IHC.

Este panorama justifica esta investigação partindo da necessidade de buscar alternativas às interfaces atuais de resultados de busca. Presume-se que o formato diagramático, associando elementos visuais e textuais, possa melhorar a percepção dos usuários em relação aos aspectos de relevância entre os documentos e a requisição de busca. Dessa forma, espera-se que este estudo possa prover à sociedade que busca acesso à informação, novas possibilidades de acesso a documentos em sistemas de recuperação de

informação.

1.2 PROBLEMA DA PESQUISA

Torna-se importante questionar se é possível e viável desenvolver uma interface para exibir os resultados de busca em um sistema de RI que possa incorporar fundamentos da IHC e da Semiótica. Pressupõe-se que a IHC discute formas de apresentação e interação, e a Semiótica procura explicar o processo de significação através do estudo dos signos e da percepção. Portanto, o problema desta pesquisa pode ser traduzido através da seguinte questão: como incrementar o processo de percepção dos usuários dos resultados de buscas, com novas combinações de recursos visuais e textuais em interfaces de Recuperação de Informação.

1.3 HIPÓTESES

Baseando-se nas ideias já apresentadas, pressupõe-se que as bases conceituais da semiótica e da IHC contribuem para o desenvolvimento de interfaces em sistemas de RI. A primeira hipótese parte da compreensão de que a associação de elementos visuais aos textuais na representação da relevância de cada documento em relação à expressão de busca, amplia as possibilidades de percepção dos usuários. Por fim, entende-se que é possível incorporar às interfaces de recuperação de informação uma visualização diagramática representando a similaridade dos documentos resultantes da busca por informação.

1.4 TESE

A partir do cenário descrito, define-se a tese desta pesquisa: as interfaces de resultados de busca em sistemas de recuperação de informação podem ser aperfeiçoadas utilizando-se recursos visuais fundamentados na semiótica e na IHC ampliando as possibilidades de percepção dos usuários.

1.5 OBJETIVOS

O objetivo geral deste trabalho é propor um modelo de interface computacional para apresentação dos resultados de busca em sistemas de recuperação de

informação, agregando elementos visuais fundamentados nas áreas de Semiótica e IHC.

Para se atingir o objetivo geral, foram estabelecidos os seguintes objetivos específicos:

- analisar o processo de recuperação de informação através de seus aspectos históricos e conceitos gerais, elencando os principais desafios a serem confrontados em relação às interfaces computacionais;
- investigar a contribuição das bases teóricas da IHC para o desenvolvimento de interfaces digitais;
- abordar o aporte filosófico da semiótica peirceana na estruturação dos elementos visuais utilizados na proposta;
- buscar tecnologias gratuitas e portáteis que viabilizem tecnicamente a elaboração da interface;
- apresentar uma interface de resultados de busca utilizando os estudos relacionados ao modelo vetorial, a IHC e a semiótica.
- realizar testes de usabilidade visando avaliar melhorias na proposta.

1.6 METODOLOGIA

Com o intuito de alcançar os objetivos apresentados, realizou-se uma pesquisa dividida em duas partes. Em um primeiro momento apresenta-se uma pesquisa bibliográfica de caráter exploratório. Em seguida foi realizada a pesquisa aplicada. Gil (2008, p.27-28) explicita que os estudos bibliográficos exploratórios devem proporcionar uma visão geral sobre os temas, explorando alguns conceitos essenciais apresentados por autores relevantes nas respectivas áreas. A partir da abordagem do problema, classifica-se este estudo como qualitativo e de natureza aplicada, segundo Prodanov e Freitas (2013). A pesquisa aplicada tem o intuito de gerar conhecimentos para aplicação prática na busca por soluções de problemas específicos. O Quadro 1 resume as atividades que compuseram o percurso metodológico utilizado na pesquisa.

Quadro 1: Percurso metodológico da pesquisa

Etapa	Temas
Pesquisa exploratória utilizando estudo bibliográfico	<ul style="list-style-type: none"> ▪ Recuperação de Informação - abordagem dos principais conceitos e modelos adotados na pesquisa e desafios na RI ▪ Interação Humano Computador - Conceitos básicos, bases teóricas e técnicas de visualização de informação ▪ Fundamentação teórica da Semiótica e o processo de percepção
Pesquisa aplicada através da proposta de interface	<ul style="list-style-type: none"> ▪ Modelagem do protótipo através de diagramas da UML ▪ Proposta de interface de resultados de busca ▪ Implementação do protótipo ▪ Discussão da interface em relação à fundamentação teórica

Fonte: Elaborado pelo autor

A pesquisa bibliográfica se estendeu durante toda a elaboração deste estudo, possibilitando a busca por conceitos necessários para o embasamento teórico. Foram analisadas referências teóricas extraídas de artigos, teses, dissertações e livros que versam sobre recuperação de informação, IHC e semiótica. Para isto, foram utilizadas as seguintes bases de dados eletrônicas: ACM, Biblioteca Digital Brasileira de Teses e Dissertações (BDTD), *Directory of Open Access Journals* (DOAJ), *Emerald Insight*, Google Acadêmico, IEEE, *Interscience Wiley*, Periódicos Capes, *Scientific Electronic Library Online* (Scielo), Scopus e *Web of Science*. As bases de dados foram selecionadas por indexarem diversos periódicos nacionais e internacionais de caráter multidisciplinar, buscando agregar estudos de várias áreas do conhecimento. Durante a busca foram utilizados os termos relacionados às três áreas (Recuperação de Informação, Semiótica e Interação Humano Computador) e suas variações no plural e na língua inglesa. Inicialmente analisou-se os títulos, assuntos, resumos e palavras-chave, selecionando e armazenando as pesquisas que apresentaram alguma relação com os objetivos desta exploração.

Os documentos retornados foram divididos em dois grupos. No primeiro, as pesquisas que serviram de fundamentação teórica. No segundo grupo, as publicações que apresentaram relação com a seção de discussão deste estudo, ou seja, que apresentaram desafios para a recuperação da informação ou aspectos importantes da IHC e da semiótica que confrontassem tais desafios. Dessa forma, a partir dos documentos selecionados foi desenvolvido o estudo sobre recuperação de informação, conceituando e listando os desafios dos SRIs. Em seguida, fundamentou-se a IHC e a semiótica, destacando os aspectos positivos que possivelmente contraponham os obstáculos na construção de interfaces de busca.

Após a fundamentação teórica, foi realizada uma pesquisa aplicada visando

propor uma interface de resultados de busca que fosse composta pelas seguintes características:

- a) Disposição dos documentos resultantes da busca em modelo visual agregando elementos visuais e textuais embasados pela IHC e semiótica;
- b) Utilização do modelo vetorial no cálculo de relevância e apresentação dos resultados tornando transparente a razão pela qual o documento foi recuperado e a relação dos documentos com a busca;
- c) Uso de estilos de interação compreensíveis baseados nos estudos de IHC;
- d) Implementação do protótipo com tecnologias gratuitas e portáteis, ou seja, acessíveis tanto no desenvolvimento de ferramentas como no seu uso.

A proposta de implementação do protótipo tem como foco responder, de forma prática, ao objetivo e questionamento da pesquisa. Procura-se agregar aos fundamentos teóricos uma ferramenta computacional que contenha aspectos descritos na literatura e, portanto, com valor científico adequado.

1.7 ESTRUTURA DA TESE

Esta introdução procura apresentar o tema, o problema, as hipóteses, a tese, os objetivos e a metodologia utilizada para abordar as ideias desta pesquisa. A sequência da tese foi estruturada ao longo de mais cinco capítulos, conforme descritos a seguir.

O Capítulo 2 versa sobre a Recuperação de Informação por ser a motivação inicial da realização deste estudo. Inicialmente são apresentados alguns aspectos históricos apresentando fatos relevantes da área. Por meio da ilustração do processo de RI, busca-se posicionar os elementos que o compõem, e também situar o foco desta pesquisa.

Com a identificação de desafios relacionados às interfaces em sistema de RI, tornou-se adequado buscar na área de IHC princípios que contribuíssem para o andamento da pesquisa. Assim, o Capítulo 3 inclui uma abordagem teórica e histórica sobre a área e apresenta técnicas de visualização de informação que visam melhorar o processo de interação dos usuários.

A semiótica é caracterizada no Capítulo 4, realçando os conceitos que embasaram a teoria da percepção peirceana. Pretende-se buscar o entendimento dos fatores

que influenciam a recuperação de informação em interfaces. São apresentadas pesquisas que se relacionaram com o desenvolvimento de interfaces utilizando-se dos princípios da área.

A proposta de interface é ilustrada no Capítulo 5. A princípio, são utilizados diagramas para a explicitação dos requisitos a serem atendidos pelo protótipo. Também são descritas as tecnologias empregadas na construção da proposta. O protótipo é detalhado através da exibição de imagens e, em seguida, é empreendida a discussão do suporte teórico das áreas de RI, IHC e semiótica na concepção da ferramenta.

As conclusões são apresentadas no sexto capítulo, tecendo considerações sobre o trabalho em relação ao problema e aos objetivos da pesquisa. Também são propostos temas para futuras pesquisas que possam dar continuidade e agregar elementos a este estudo.

2 RECUPERAÇÃO DE INFORMAÇÃO

Uma das bases conceituais deste estudo é a Recuperação de Informação (RI). Neste capítulo serão apresentados alguns conceitos gerais, realçando principalmente os elementos que servem como alicerce para as ideias descritas no decorrer da pesquisa. A área pode ser estudada a partir de duas perspectivas, ou seja, uma focada nos sistemas computacionais, e outra no usuário. Os estudos centrados em sistemas computacionais buscam principalmente a melhoria de desempenho nas consultas, otimização dos algoritmos de ranqueamento e desenvolvimento de interfaces digitais que possam melhorar a exibição dos resultados. Nas pesquisas focadas nos usuários destacam-se fatores referentes ao comportamento informacional, buscando entender suas necessidades e o impacto delas nos sistemas de RI. (BAEZA-YATES; RIBEIRO-NETO, 2013)

A seguir, são destacados alguns aspectos históricos que nortearam as pesquisas sobre recuperação de informação e também são apresentadas algumas contribuições de autores que servem de referência no que diz respeito ao desenvolvimento da área. Nas próximas seções, algumas concepções iniciais sobre os Sistemas de Recuperação de Informação são apresentadas, no intuito de revelar as motivações e desafios encontrados nas pesquisas sobre o tema.

2.1 ASPECTOS HISTÓRICOS

O termo *Information Retrieval* (Recuperação de Informação - RI) foi apresentado, originalmente, por Mooers (1951). O autor descreve a RI como um processo pelo qual os usuários são capazes de converter as suas necessidades de informação em uma lista real de citações para documentos armazenados que contêm informações úteis para ele. Baeza-Yates e Ribeiro-Neto (2013) apresentam alguns pesquisadores que contribuíram historicamente com trabalhos relevantes no desenvolvimento da temática. Dentre eles destaca-se Gerard Salton, que a partir dos anos 1960 colaborou para dar forma à área, além de publicar o primeiro livro dedicado a RI, em 1968.

Durante os anos subsequentes, as pesquisas voltadas para o tema percorreram conceitos fundamentais, abordando os processos de recuperação, tecnologias de ranqueamento, métricas e os modelos de Recuperação de Informação. Conforme descrito por Ferneda (2013), ainda na década de 1960 Maron e Kuhns apresentaram os princípios básicos do modelo

probabilístico e Salton realizou experimentos em seu projeto SMART que possibilitou a publicação de diversos artigos científicos, além do desenvolvimento de técnicas computacionais para a RI. Na década de 1970, Cornelis Joost Van Rijsbergen publicou o livro *Information Retrieval* em que apresentou um panorama da área, e descreveu o modelo probabilístico (RIJSBERGEN, 1979). Em 1983, outro livro clássico foi lançado por Salton e McGill, *Introduction to Modern Information Retrieval* (SALTON; MCGILL, 1983), no qual é apresentado o modelo vetorial.

Conforme exposto, os modelos de recuperação de informação que fizeram parte dos estudos iniciais da área, podem ser divididos em clássicos e estruturados. São três os modelos clássicos de recuperação: o modelo booleano, o modelo vetorial e o modelo probabilístico. Souza (2006, p. 166) descreve que:

[..] nos modelos clássicos cada documento é descrito por um conjunto de palavras-chave representativas – também chamadas de termos de indexação – que busca representar o assunto do documento e sumarizar seu conteúdo de forma significativa. Nos modelos estruturados, podem-se especificar, além das palavras-chave, algumas informações acerca da estrutura do texto (como seções a serem pesquisadas, fontes de letras, proximidade das palavras, entre outras informações.).

Seguindo o percurso histórico, o advento da *World Wide Web*, a partir da década de 1990, ocasionou o crescimento da quantidade de documentos disponíveis, pois proporcionou aos usuários criarem, disponibilizarem e pesquisarem documentos produzidos no mundo digital. Neste panorama, os estudos em RI se tornaram imprescindíveis, pois possibilitam estudar e desenvolver técnicas que objetivam o acesso adequado a informações contidas nos mais variados tipos de documentos. Embora sucinto, este histórico apresenta contribuições de alguns pesquisadores basilares para a afirmação da RI como área de estudo. Seguindo este percurso, a seguir serão apresentadas algumas concepções iniciais da Recuperação de Informação, elencando quais pontos de vista esta pesquisa pode seguir.

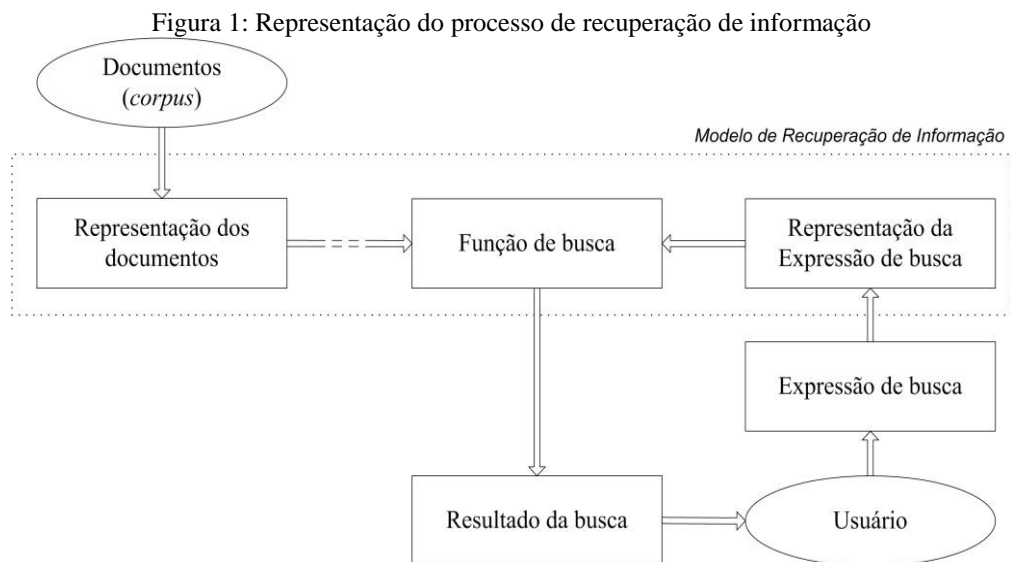
2.2 O PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO

Para iniciar a discussão sobre Sistemas de Recuperação de Informação, Salton e McGill (1983) caracterizam os SRIs como sistemas que trabalham com as tarefas de representação, armazenamento, organização e acesso aos itens de informação. Souza (2006) argumenta que um Sistema de Recuperação de Informação apenas informa sobre a existência de documentos relacionados à sua requisição, ficando a cargo do usuário filtrar os documentos realmente relevantes. Desta forma, a RI é tema de diversas pesquisas que têm

como objetivo a melhoria da eficiência deste processo. Na visão de Ferneda (2013, p. 12):

A recuperação de informação envolve, por um lado, um acervo documental composto de itens informacionais. Por outro lado, temos pessoas, seres humanos com as mais variadas necessidades de informação que buscam por documentos relevantes para satisfazer tais necessidades. Recuperar informação implica, portanto, em operar seletivamente um estoque de informação, o que envolve processos cognitivos difíceis de serem formalizados. A utilização de recursos computacionais nessa tarefa parte de inevitáveis simplificações teóricas e de adequações de conceitos subjetivos tais como “relevância” e “necessidade de informação”, além do próprio conceito de informação.

Para ilustrar o processo de recuperação de informação, a Figura 1 apresenta uma representação simplificada.



Fonte: Ferneda (2013, p. 48)

Observa-se que para recuperar informação é necessário que exista uma relação entre a representação dos documentos e a expressão de busca do usuário. A partir desta relação, identifica-se no acervo documental quais atendem às necessidades de informação do usuário. As próximas seções caracterizam os elementos apresentados na Figura 1.

2.2.1 Representação de documentos

Entende-se que os documentos para serem recuperados devem ser representados. Sales (2017, p. 11) discute a representação documental e esclarece que "representar o objeto informacional (documento) é encontrar seus substitutos sintéticos, os

elementos que o (re)apresentarão", ou seja, representar um documento é apresentá-lo novamente, utilizando elementos com potencial de fornecer pontos de acesso à informação nele contida. Inicialmente torna-se pertinente o entendimento de que o conceito de documento a ser considerado refere-se à sua capacidade informativa.

Ortega e Lara (2010) realizaram uma análise de documento como objeto informacional, traçando um percurso histórico dos conceitos de documento e documentação. No artigo, as autoras apresentam conceitos desde os primeiros documentalistas, Paul Otlet e Suzanne Briet, destacando que Briet foi umas das principais continuadoras das ideias de Otlet. Neste cenário, o conceito configura documento como todo signo indicial (ou índice) concreto ou simbólico, preservado ou registrado para fins de representação, de reconstituição ou de prova de um fenômeno físico ou intelectual.

Dentro da mesma temática, utiliza-se uma das abordagens de Buckland (1991) sobre informação que se refere ao conceito de "informação como coisa" (*information-as-thing*) onde atribui-se a capacidade informativa a um objeto. O autor considera que nos acervos documentais, a informação está relacionada ao objeto que a contém, assim o termo documento pode ser entendido como algo informativo, como por exemplo: livros, textos e obras de arte.

Devido ao fato desta tese estar voltada aos estudos de interfaces de Recuperação de Informação, e por consequência o acesso à informações neste meio, torna-se relevante destacar aspectos sobre a noção de "documento digital". Buckland (1998) salienta que as concepções de Otlet e Briet possibilitam inferir que a noção de "documento" independe do meio utilizado e contempla não somente formas físicas. Neste sentido, as definições baseadas em forma, formato e meio estão inerentemente ligadas a uma abordagem funcional, que considera a função informacional do objeto, embora possuam diferentes capacidades e restrições.

A partir do mesmo panorama, Siqueira (2012) empreende uma abordagem terminológica sobre a noção de documento digital, apresentando diversos elementos históricos relativos ao conceito de documento e sua evolução. Também discute uma análise terminológica optando por autores como Otlet e Briet, os quais fazem parte da linha de estudo considerada como berço da tradição documentária. A autora constata que a noção de documento digital está conectada à noção de documento e também às concepções de informação. Desse modo, a função do documento é representar uma mensagem com capacidade de ser utilizada como informação. Enfim, o documento digital independe de suas particularidades, sendo uma manifestação do documento em um meio estabelecido.

Concluindo, torna-se relevante "destacar a importância do documento digital como um objeto de interatividade, por propiciar maior rapidez e fluidez na recuperação e comunicação informacional" (SIQUEIRA, 2012, p.137).

O principal objetivo de representar um documento é o de identificar e caracterizar, sinteticamente, o conteúdo nele existente. Essa representação deve também estabelecer pontos de acesso para que os sistemas de busca de informação possam recuperá-lo. Fernalda (2013, p. 48) salienta que a "tarefa de representar os documentos é feita em um tempo anterior à execução de qualquer busca". Novellino (1996, p. 38) destaca:

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.

Importante considerar que em toda representação existe uma perda, mas que é justificável, entendendo a necessidade de redução com o fim de limitar apenas o que é significativo do conteúdo do documento para o processo de recuperação. Dessa forma, o processo de recuperação torna-se viável, ou seja, a cada solicitação de busca são acessadas apenas as representações dos documentos, o que agiliza a geração dos resultados, pois não existe a necessidade de rastrear todo o conteúdo de cada documento do acervo .

A representação pode ser dividida entre descritiva e temática, Maimone, Silveira e Tálamo (2011) explicam a representação descritiva como a descrição bibliográfica através de características específicas, padronizando pontos de acesso, como por exemplo, autor ou título do documento. Já a representação temática atenta-se a representar os assuntos com o intuito de aproximar documentos com temáticas semelhantes. Necessário salientar que as duas formas se complementam para facilitar o acesso eficiente no processo de RI. Esses tipos de representação indicam um importante fator de relação entre usuários e documentos. A representação temática é uma das formas de indexação a ser considerada quando se tratar de sistemas informatizados de recuperação de informação; ela também possibilita o uso de tecnologias computacionais para automatizar o processo. Predizendo futuras discussões desta pesquisa, a seguir será exposta a indexação, em especial o processo automatizado, tendo em vista a utilização dos pontos de acesso (metadados) de documentos no processo de desenvolvimento da interface proposta.

Borges (2009) descreve que a indexação é a tarefa de representar um documento de forma sintética e pode ser realizada através de uma operação intelectual humana ou de forma automática. O modelo intelectual aposta em estabelecer a representação temática próxima ao conteúdo do documento através da capacidade de abstração profissional. Já a indexação automática se apresenta como alternativa à morosidade do processo realizado pelos humanos frente a grandes volumes de documentos. Na visão de Cintra (1983), o objetivo do indexador é eleger palavras-chave ou descritores que expressem o conteúdo de um documento de forma a possibilitar a sua recuperação. A automatização busca utilizar a tecnologia com o objetivo de agilizar o processo. Os termos atribuídos no processo de indexação têm a função de atuar como pontos de acesso no processo de localização e recuperação dos documentos.

De acordo com Lancaster (2004), a indexação pode ser realizada de duas formas, ou seja, por extração ou por atribuição. A indexação por extração restringe o uso de termos ao contexto do próprio documento, onde aplicam-se critérios institucionais e pessoais na seleção dos termos que farão a representação do conteúdo. No caso da indexação por atribuição, utiliza-se algum elemento externo ao documento no processo de indexação. São predeterminados conjuntos de termos e os conjuntos podem ser listas simples de termos controlados ou até mesmo estruturas mais complexas como tesouros e ontologias, a partir dessa lista o indexador seleciona os termos mais adequados.

Da mesma forma do processo de indexação realizado por humanos, Lancaster (2004) também divide o processo de indexação automática em indexação por extração automática e indexação por atribuição automática. Com o intuito de limitar o escopo desta pesquisa, torna-se necessário direcionar a discussão em particular ao processo de indexação por extração automática. Esta opção é pertinente, pois viabiliza o desenvolvimento da interface proposta através do uso de *softwares* e algoritmos já disponíveis que utilizam a indexação por extração automática. Além disso, esta ideia atende aos objetivos de oferecer pontos de acesso aos documentos que serão expostos aos usuários.

Conforme apresentado por Borges (2009), a indexação automática iniciou-se no final da década de 1950, quando Hans Peter Luhn propôs a ideia de que a essência do conteúdo de um documento poderia ser extraída através do vocabulário presente nele, constituindo, assim, um forma adequada de recuperá-lo. O método de Luhn para atribuição de índices procurava inicialmente excluir do texto palavras de pouco valor semântico. Robredo (1982) descreve que durante o processo de indexação automática é realizada a comparação das palavras do texto com uma lista previamente determinada de palavras vazias de

significado, tais como artigos, preposições e conjunções. Essas palavras são eliminadas e as restantes são consideradas como palavras representativas para aquele conteúdo. Mesmo utilizando esta técnica, a lista ainda gera muitas entradas temáticas de índice. Neste sentido, ainda seriam necessárias outras formas de filtro.

Ainda segundo Robredo (1982), duas situações podem ser observadas. A primeira diz respeito a utilizar apenas as raízes das palavras eliminando prefixos e sufixos. Neste sentido, assume-se que palavras com o mesmo radical são semanticamente relacionadas, tendo o mesmo significado. Como exemplo o radical "relac" poderia ser atribuído às suas variações como "relacionada", "relacionado", "relaciona", "relacionam-se", entre outras. A segunda situação seria a identificação de expressões significativas formadas por mais de uma palavra, ou seja, identificar ligações entre palavras que formariam expressões compostas. Infere-se que termos compostos agregam maior valor semântico à representação do conteúdo de um documento, potencializando a recuperação de informação, visto que são menos ambíguos do que as palavras que os compõem.

Importante distinguir os conceitos de termo simples, composto e complexo. Conforme considerado por Barros (2007), termos simples são as unidades terminológicas compostas por um único lexema (unidade mínima distintiva do sistema semântico de uma língua que reúne todas as flexões de uma mesma palavra), como exemplo: livro, caderno. Já os compostos são formados por dois ou mais lexemas em situação de não autonomia representada graficamente com a utilização de hífen, como por exemplo: pós-graduação. Por fim, os termos complexos são formados por composições sintagmáticas não ligadas por hífen, como: Ciência da Informação. Também no que diz respeito à combinação de palavras, Kuramoto (1995, p.3) apresenta o conceito de sintagma nominal como "a menor parte do discurso portadora de informação" e, sendo assim, a combinação dos elementos formam um conceito. O autor exemplifica a situação através da expressão "livro de bolso", em que as palavras isoladamente são palavras da língua, enquanto a sua combinação faz referência a uma classe de objetos que traz maior especificidade na compreensão das coisas.

Após a geração do conjunto dos termos de indexação, que podem ser palavras, radicais ou termos compostos, entende-se necessário diferenciar o grau de importância de cada termo na tarefa de representar o conteúdo informacional de um documento. Diante disso, Losee (2001) salienta que Luhn fundamentou-se na proposta do princípio de mínimo esforço formulada pelo linguista Kingsley Zipf (1902-1950). Este princípio considera que a partir da relação das palavras em ordem decrescente de frequência, a posição de cada palavra, multiplicada por sua frequência, resulta um valor praticamente

constante. Dessa forma, não seriam as palavras mais frequentes eleitas como mais adequadas e nem todos os termos de indexação contribuem igualmente na representação do conteúdo informacional do texto. Na proposta de Luhn, termos mais significativos são os de frequência média, observando que as palavras muito frequentes teriam baixo poder para representar o conteúdo informacional do documento, e as palavras com frequência muito baixa seriam pouco significativas na representação do documento.

As definições acima mostram o processo de indexação automática por extração. Em resumo temos as seguintes etapas:

- a) reconhecer as palavras do texto;
- b) eliminar palavras que são insignificantes para representar o conteúdo do documento;
- c) restringir as palavras a uma forma normalizada ou à sua raiz;
- d) buscar e estabelecer termos compostos e complexos como termos de indexação;
- e) determinar os pesos de cada termo.

Concluindo, a questão da indexação foi apresentada brevemente nesta seção com o intuito de auxiliar e fundamentar o processo de representação de documentos para sistemas automatizados. Os conceitos de indexação servem como base para os modelos de recuperação de informação; esses modelos proporcionam a relação entre a disponibilização dos documentos e os anseios dos usuários que buscam por informações nos acervos documentais. Neste sentido antes de discorrer sobre os modelos, será analisado o outro extremo do processo, ou seja, os usuários na RI.

2.2.2 Usuários na Recuperação de Informação

Diante da atual diversidade de indivíduos que utilizam sistemas de RI na busca por informação, a tarefa de conceituar o termo usuário pode parecer imprecisa, isto devido às variáveis consideradas nos estudos relacionados a usuários. De forma mais generalista, Dantas, Silva e Souza (2013) caracterizam usuários como indivíduos que utilizam serviços de busca por informação, com o objetivo de preencher suas lacunas informacionais agregando novos conhecimentos ou ampliando-os para constituir novos comportamentos na sociedade.

No exame da relação dos usuários com a RI, identificam-se diversas peculiaridades. No contexto da Ciência da Informação o termo Recuperação de Informação

significa a operação pela qual se selecionam documentos a partir do acervo, em função da demanda do usuário. A Recuperação de Informação também consiste no fornecimento, a partir de uma demanda definida pelo usuário, dos elementos de informação documentária correspondentes (FERNEDA, 2003, p. 14). A partir disso, é possível visualizar a pertinência dos estudos de usuário e de comportamento informacional. Seguindo a mesma tendência, Chowdhury (2014, p. 4), salienta: "[...] o usuário é o ponto focal de todos os sistemas de recuperação de informação, pois o único objetivo de qualquer sistema de armazenamento e RI é transferir informação da fonte (bases de dados) ao usuário". No que diz respeito ao início dos estudos de usuários na Ciência da Informação, Rolim e Cendón (2013, p. 1, grifo nosso) destacam:

Os primeiros estudos de usuários na ciência da informação foram desenvolvidos a partir da percepção da **necessidade da informação** do público comum da biblioteca pública ou do uso das fontes de informação de cientistas. Em ambos os casos, os estudos evoluíram e as novas correntes refletiram em uma tentativa de transferir o foco do sistema para o usuário, usuário considerado primeiramente um ser passivo diante da informação, mas que então **passa a ser visto como um ser ativo**, que constrói perguntas, respostas e caminhos para suas demandas de informação.

A partir dessa afirmação destacam-se dois pontos: a proatividade do usuário e a concepção de necessidade de informação. Nessa perspectiva, verifica-se uma atuação dinâmica dos usuários diante de fontes de informação. Rabello (2013, p. 181) salienta que "o conceito de usuário passou a contemplar, em um só tempo, o sujeito que usa, produz e se apropria de informação e de conhecimento". No que tange à expressão "necessidade de informação", embora subjetiva, essa concepção pode ser abordada a partir de diferentes pontos de vista. Chowdhury (2014) considera necessidade informacional um conceito vago que pode ocorrer quando o usuário assume ter conhecimento insatisfatório para realizar algo ou solucionar divergências em determinada área. O autor elenca alguns pontos a serem ponderados:

- a) Necessidade de informação é um conceito relativo; ela depende de diversos fatores e não permanece constante;
- b) Ela muda sobre um período de tempo;
- c) Varia de pessoa para pessoa, de profissão para profissão, de área para área, organização para organização, e assim por diante;
- d) As necessidades informacionais das pessoas são em grande medida dependentes de seu ambiente; por exemplo, as necessidades daqueles que estão no ambiente acadêmico são diferentes daqueles num ambiente

- industrial, de negócios, ou administrativo;
- e) Mensurar (quantificar) a necessidade de informação é difícil;
 - f) Frequentemente, essa necessidade permanece inexpressável ou insuficientemente expressa;
 - g) A necessidade de informação frequentemente muda ao se receber alguma informação.

A abordagem expõe aspectos que diferem as necessidades informacionais, dependendo das características do indivíduo, contudo entende que existindo a necessidade de resolução de um problema, os usuários buscam informação para tentar compreender e solucionar tal situação. A condição de dúvida é característica do início do processo de busca de informação, e conforme descrito acima, as necessidades nem sempre são expressas de maneira satisfatória. Entretanto, quando o indivíduo explora fontes informacionais, a sua condição de conhecimento evolui, permitindo formular buscas mais adequadas. A constituição do conhecimento inicia-se na identificação de uma necessidade que impulsiona o processo de busca. Dessa forma, a utilização racional da informação proporciona a geração do conhecimento a ser aplicado à prática pessoal, profissional, social ou acadêmica.

Wilson (2005) verificou a evolução de seus modelos de comportamento informacional. O autor sugeriu que a "necessidade de informação" não seria o conceito apropriado para o propósito de pesquisa. Esclareceu, ainda, que o termo "comportamento de busca por informação" pode ser adotado, já que é observável, enquanto as necessidades não o são. Assim, verifica-se a importância dos estudos de comportamento informacional no processo de recuperação de informação. Wilson (2000, p. 49, tradução nossa) define "comportamento informacional" como

[..] a totalidade do comportamento humano em relação às fontes e canais de informação incluindo busca e uso ativo e passivo da informação. Assim, ele inclui a comunicação face a face, bem como a recepção passiva da informação, por exemplo, assistir propaganda na televisão, sem qualquer intenção de agir sobre a informação dada; o comportamento de busca de informação é direcionado para a busca da informação como consequência de uma necessidade, para satisfazer um objetivo. Durante a busca, o indivíduo pode interagir com sistemas de informação manual, um jornal ou uma biblioteca, ou com sistemas automatizados como, por exemplo, a web.

A definição acima apresenta o conceito de "comportamento informacional" na sua totalidade, porém, nesta pesquisa, limita-se ao comportamento de busca e recuperação da informação num sistema automatizado. Na análise da relação que se estabelece entre os

sistemas computacionais de RI e o comportamento informacional dos usuários, percebe-se que são apresentadas peculiaridades que podem ser explicadas a partir de padrões observados em revisão de literatura. Neste sentido, as investigações sobre comportamento informacional e comportamento de busca, geraram modelos. A seguir serão descritos alguns modelos frequentemente apresentados nos estudos da área.

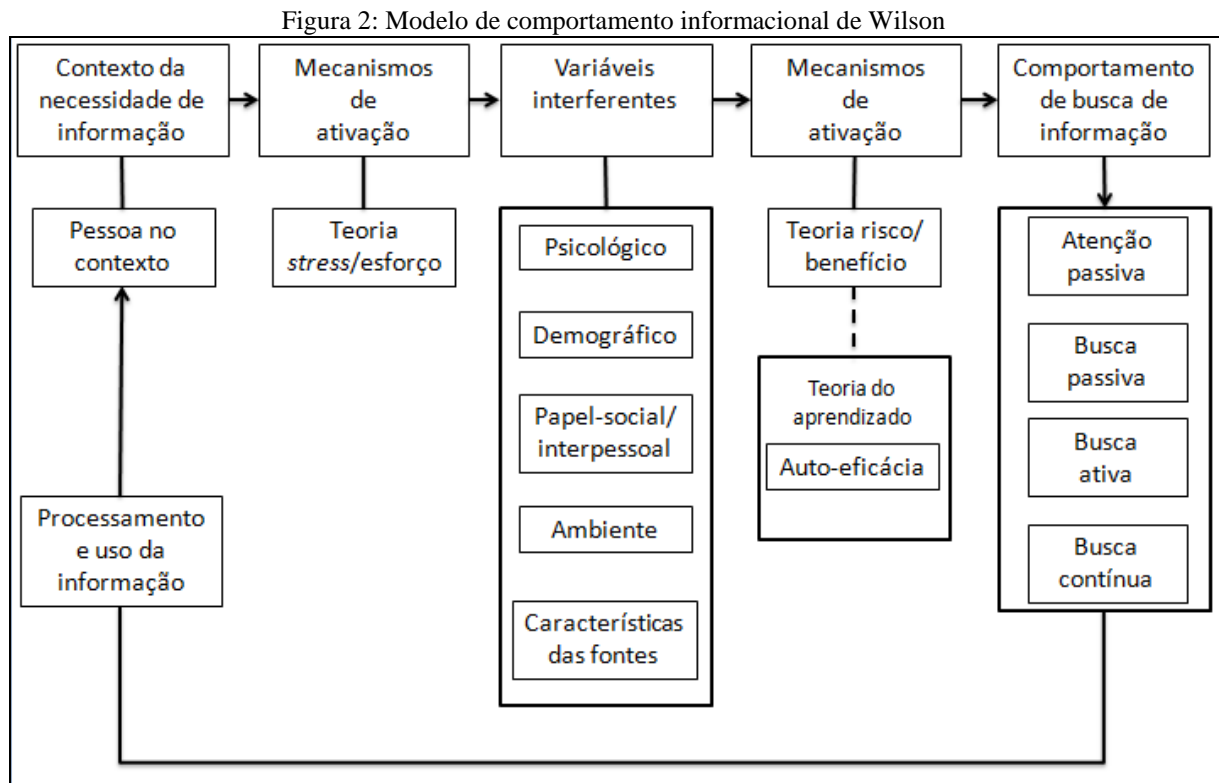
2.2.2.1 Modelos de comportamento de busca

Segundo Martínez-Silveira e Oddone (2007), a partir dos anos 1980 os estudos sobre comportamento e necessidades informacionais passaram a contemplar a perspectiva dos usuários, em detrimento à ênfase dada anteriormente aos sistemas em si. Na mesma época as pesquisas passaram a valorizar técnicas qualitativas em substituição às quantitativas.

Belkin, Oddy e Brooks (1982) apresentam o conceito dos Estados Anômalos do Conhecimento (*Anomalous States of Knowledge – ASK*) para abordar quando um indivíduo identifica ausência de determinada informação e, por consequência, reconhece a necessidade de busca por novas informações. Rolim e Cendón (2013) destacam que "estado anômalo" é a percepção desse estado inicial com ausência de informação, incertezas e incoerências. Ao buscar informação, o estado de conhecimento do usuário é alterado e ele avalia a necessidade de realizar novas buscas; neste caso o usuário pode reavaliar suas fontes, mudar a forma de pesquisa e finalizar suas buscas, de acordo com sua motivação.

Wilson formulou, em 1981, um modelo focado em padrões do comportamento informacional humano, tendo em vista as necessidades informacionais dos usuários. Este modelo foi revisado em 1996 e é apresentado na Figura 2. Furtado e Alcará (2015) apresentam a primeira versão (1981) a qual definiu as necessidades de informação como psicológicas, afetivas ou cognitivas e as relações dos indivíduos na sociedade e ao ambiente em que está inserido, considerando aspectos políticos, econômicos ou tecnológicos exercidos por ele. A partir das necessidades de informação, é realizada uma busca ativa a partir de duas proposições: na primeira, a informação é considerada uma necessidade secundária originada a partir de necessidades primárias; a segunda proposição enfatiza a existência de possíveis barreiras enfrentadas pelos indivíduos na busca por informação. Wilson aperfeiçoou seu modelo em 1996, incluindo alguns conceitos como mecanismos de ativação, caráter cíclico da busca, importância do contexto e categorização de variáveis intervenientes, envolvidas com os aspectos individual, social e ambiental do indivíduo. Nesta

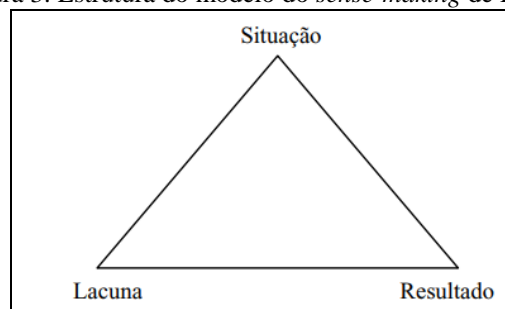
reorganização, Wilson utilizou o modelo *Sense Making* para incluir o conceito descrito por Dervin (1983) como "lacuna", ou seja, um estágio entre a pessoa e a consciência da necessidade de informação.



Fonte: (WILSON, 2000, p. 53)

Dervin desenvolveu a teoria conhecida como *sense-making* (Figura 3), formada por três elementos: situação, lacuna e resultado. A situação, em tempo e espaço, seria o contexto no qual surge o problema informacional; a lacuna (*gap*) seria a distância entre a situação contextual e a situação desejada (incerteza) e, por fim, o resultado representa a consequência do processo de *sense-making* (DERVIN, 1983).

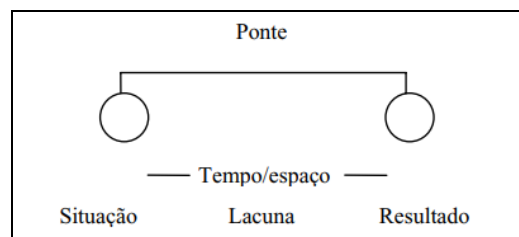
Figura 3: Estrutura do modelo do *sense-making* de Dervin



Fonte: (DERVIN, 1983, p. 9)

Conforme descrito por Martínez-Silveira e Oddone (2007), o modelo de Dervin é de natureza cognitiva e considera a necessidade de informação como subjetiva, situacional e holística. Neste sentido, o usuário é considerado ponto central do processo, além de sujeito ativo. Conforme apresentado na Figura 4, Dervin exemplificou a "ponte" como meio de preencher a lacuna entre a situação e o resultado. Com o objetivo de preencher suas lacunas informacionais os indivíduos estudam, pesquisam ou conversam com outras pessoas adquirindo, gradativamente, informações que satisfaçam suas necessidades. Cada etapa na aquisição de conhecimento representa uma ampliação de degraus na experiência adquirida do indivíduo. Os degraus relacionados à definição e ao fechamento da lacuna constituem estratégias cognitivas necessárias à obtenção de respostas, algo complexo e dependente do indivíduo e da situação. Neste percurso, o indivíduo precisa de esforço para se perceber, perceber o meio ambiente e ir efetuando os ajustes necessários (MARTÍNEZ-SILVEIRA; ODDONE, 2007).

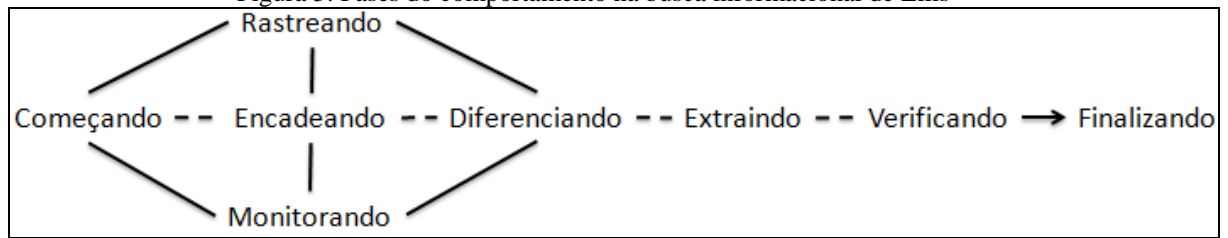
Figura 4: Metáfora do modelo *sense-making* de Dervin



Fonte: (MARTÍNEZ-SILVEIRA; ODDONE, 2007, p. 123)

David Ellis desenvolveu, em 1989, um modelo que apresentava uma sequência de categorias na tarefa de busca informacional que envolvia: começar (atividades de início da busca); encadear (prosseguir a busca); browsing (busca semidirigida em locais potenciais de busca); diferenciar (filtrar e selecionar); monitorar (continuar revendo as fontes identificadas como essenciais); extrair (trabalhar sistematicamente com as fontes de interesse); verificar (conferir a veracidade das informações). Para Wilson (1999), a relevância do modelo de Ellis consiste no fato de ser resultado de uma pesquisa empírica, além de ter sido testado em diversos estudos. Ellis (1989) salienta que a ideia de usar os aspectos comportamentais das atividades de busca de informações dos usuários para embasar projeto de sistemas de recuperação de informações é totalmente aplicável e poderia desempenhar um papel mais proeminente no design de sistemas de recuperação de informação, baseados em computador.

Figura 5: Fases do comportamento na busca informacional de Ellis



Fonte: (MARTÍNEZ-SILVEIRA; ODDONE, 2007, p. 124)

Choo, Detlor e Turnbull (1998) exemplificam algumas dessas atividades no modelo de Ellis, a partir de buscas utilizando um navegador de internet:

- a) A atividade "começando" indica quando o indivíduo inicia a busca em algumas páginas;
- b) a tarefa "encadeando" ocorre ao navegar por hiperlinks para recursos relacionados;
- c) "rastreando" diz respeito a percorrer as páginas e fontes;
- d) "diferenciando" ocorre quando se seleciona como favoritos algumas fontes para futuras visitas;
- e) a opção "monitorando" considera a ação de assinalar serviços de alerta por correio eletrônico para receber informações;
- f) e a tarefa de "extrair" seria pesquisar uma fonte específica sobre todas as informações necessitadas ou sobre um tópico em particular (extrair).

O modelo *Information Search Process* (ISP) foi elaborado por Carol C. Kuhlthau, em 1991. Este modelo relaciona as necessidades cognitivas com as reações emocionais no processo de busca por informação. Martínez-Silveira e Oddone (2007) descrevem seis estágios referentes ao nível de incerteza no processo de busca por informação. Estes seis níveis são divididos em três campos, a saber, o emocional, cognitivo e físico:

- a) Estágio de iniciação: quando há o reconhecimento da necessidade de informação;
- b) Estágio de seleção: trabalho de delimitar o campo ou tema de investigação;
- c) Estágio de exploração: investigação dos documentos acerca do tema, levando a uma expansão do tema geral (por exemplo, a leitura das fontes secundárias);
- d) Estágio de formulação: quando ocorre o estabelecimento de foco ou

- perspectiva do problema;
- e) Estágio de coleta: coleta por meio da interação com sistemas e serviços de informação para a reunião de informações;
 - f) Estágio de apresentação: o fim da busca e solução do problema.

A partir das fases apresentadas no processo, ocorre a construção do conhecimento e significado. Rolim e Cendón (2013) explicam que o processo de busca é influenciado pelo foco de interesse do usuário, e o estabelecimento deste foco ocorre a partir da interpretação das informações existentes. Dessa forma, a natureza da informação descoberta pode afetar o usuário, visto que informações redundantes podem gerar insatisfação, assim como uma nova informação pode exigir uma readequação de conhecimento ainda não disponível, provocando ansiedade. O posicionamento do usuário influencia no resultado de busca, pois o processo depende de atitudes pessoais, assim o interesse do indivíduo aumenta à medida que o foco é estabelecido e sua pesquisa avança.

Em 1992, Ingwersen propôs um modelo focado no reconhecimento dos processos cognitivos dos elementos existentes no processamento de informação. Pavão (2014) elencou os principais pontos do modelo de Ingwersen, como segue:

- a) Interação com o sistema de recuperação é um conjunto de processos de cognição que ocorre durante todo o processamento da informação e os usuários se envolvem em uma variedade de modelos cognitivos;
- b) Os usuários interagem não só com os sistemas de recuperação, mas também com os objetos de informação como, por exemplo, textos, que são estruturas cognitivas consideradas como um espaço de informação;
- c) O espaço cognitivo do usuário é um conjunto de elementos de causalidade, onde predominam o contexto cognitivo e situacional do usuário;
- d) As interações ocorrem em diferentes níveis e são de tipos diferentes;
- e) O processo é altamente dinâmico.

No modelo de Ingwersen, de 1996, Pavão (2014) complementa que a recuperação da informação se formata a partir de transformações concebidas por uma variedade de atores com diferentes perfis cognitivos. Esses indivíduos podem ser desde desenvolvedores de sistemas, indexadores, autores textuais ou usuários de determinados contextos sociais. A compreensão deste modelo exige compreender estruturas cognitivas

como ocorrências da cognição humana, reflexões ou ideias.

Conforme visto, geralmente os modelos que analisam o comportamento informacional apresentam o processo de busca constituído por fases ou etapas, às vezes interativas. Para Venâncio e Nassif (2008), o princípio da busca por informação é provocado pela constatação da existência de uma "lacuna" informacional, ou seja, quando se identifica que o estado cognitivo é incapaz para resolução de um problema específico. A partir dos estudos descritos e as contribuições teóricas, identificou-se as diversas abordagens utilizadas para atender aos mais variados objetivos de pesquisa e as possíveis maneiras de tratamento da temática deste estudo. A seguir, relata-se algumas pesquisas com a finalidade de coletar elementos que possibilitem compreender melhor o que aqui foi exposto sobre comportamento na busca por informação e sua relação com a RI

2.2.2.2 Estudos sobre comportamento do usuário na recuperação de informação

Após a apresentação dos modelos de comportamento, será apresentado nesta seção estudos que os relacionam com a RI. Pavão (2014, p. 85) apresentou a comparação de três modelos de comportamento informacional. Na visão da autora, "cada modelo representa fenômenos diferentes: Krikelas o do comportamento na busca de informação, Kuhlthau o processo da busca de informação e Wilson, num campo mais genérico, o comportamento informacional". Os modelos de Kuhlthau e Wilson baseiam-se no princípio cognitivista, ao passo que Krikelas atenta-se ao comportamentalista. Além disso, esses modelos fundamentam estudos capazes de proporcionar características relevantes na construção de sistemas que aspiram estabelecer uma relação mais satisfatória com o usuário.

Considerando os princípios cognitivistas, Lima (2003) destaca a recuperação de informação e os estudos de interação humano-computador como áreas aproximadas pela intersecção entre a ciência da informação e a ciência cognitiva. A autora defende ser essencial a elaboração de interfaces de fácil interação em sistemas de RI e para isso sugere reformular essas interfaces a partir da análise de características cognitivas no comportamento dos usuários que buscam informação.

Ainda a partir da abordagem cognitiva, Beheshti et al. (2014) investigaram o comportamento informacional de estudantes (baseado no modelo de Kuhlthau). Os autores destacam o conceito da temporalidade como um princípio fundamental da abordagem cognitiva. Nesta perspectiva, as necessidades de informação do usuário mudam ao longo do tempo e a evolução de certos conceitos definem se determinado conhecimento é considerado

relevante ou irrelevante. Neste modelo pressupõe-se que no início do processo o indivíduo ainda não tem conhecimento algum sobre o tópico a ser pesquisado, no entanto, em uma primeira pesquisa, o resultado dispõe de elementos que possibilitam formular uma nova consulta mais elaborada. Neste caso a pessoa adquire informações pouco a pouco, evoluindo e reformulando suas consultas através dos resultados recebidos, até o momento em que satisfaça suas necessidades informacionais.

Outra característica pertinente foi investigada por Pajic (2014). O estudo apresentou a utilização de buscas visuais no processo de recuperação de informação. Dentre vários fatores destacou a possibilidade de troca de informações entre a psicologia e a ciência da informação, salientando a importância de uma compreensão psicológica no processo de recuperação de informação como fundamental na resposta às necessidades dos usuários. Em sua pesquisa, o autor sugeriu a utilização de elementos gráficos como imagens e diagramas para representação dos resultados de busca. Essas representações devem agir como uma forma de ferramenta de externalização cognitiva, com o objetivo de ampliar o conhecimento do usuário, aumentando seus recursos cognitivos, reduzindo o tempo de busca, melhorando o reconhecimento de padrões, facilitando a manipulação de dados e, portanto, possibilitando a exploração holística da informação.

Visando seguir novos caminhos nos estudos sobre comportamento informacional, Huang (2007) utiliza a semiótica para destacar princípios de aquisição de conhecimento e compreensão da informação pelos usuários. Para o autor a semiótica fornece uma estrutura fundamental para analisar como as pessoas interagem com as várias mídias na busca por informação. Nesta visão, a informação é criada pelos usuários estabelecendo a conexão entre o ato de compreender símbolos ou objetos representativos e a produção de significados em um ambiente comunicativo. Especialmente ao reconhecer a concepção de informação como símbolo, devem ser analisadas três dimensões e as relações entre elas: a dimensão pragmática, a dimensão semântica e a dimensão sintática. Segundo a semiótica, a relação entre as três dimensões visa explicar a maneira como os indivíduos necessitam e interagem com a informação em diferentes contextos, a fim de obter conhecimento que atendam a suas necessidades informacionais. O autor conclui que analisar as relações entre o usuário, a representação e o significado da informação pode estabelecer uma abordagem mais holística, que não exclui o material ou o aspecto cognitivo para o estudo do comportamento informacional na pesquisa em ciência da informação.

A partir das pesquisas elencadas e as contribuições teóricas dos autores, foi possível identificar as diversas abordagens utilizadas para atender aos objetivos de pesquisa e

as possíveis maneiras de tratamento da temática deste estudo. As referências citadas possibilitam situar os estudos de usuário no processo de RI. Segundo esse panorama, é inevitável que os sistemas de informação atendam às necessidades de informação dos usuários, levando-se em conta os seus padrões de comportamento na busca da informação.

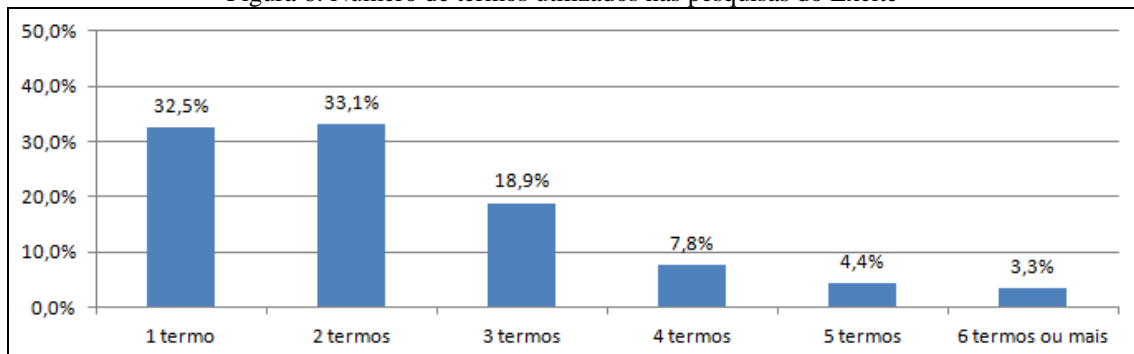
2.2.2.3 Expressão de busca e sua representação

Para expressar suas necessidades informacionais é necessário que o usuário elabore uma expressão de busca. Esta especificação pode utilizar a linguagem natural ou uma linguagem controlada, dependendo das ferramentas disponíveis no sistema. A principal dificuldade é presumir quais os termos utilizados para representar os documentos que satisfaçam sua necessidade. Em geral o usuário elabora e expressa sua necessidade informacional utilizando linguagens textuais, e também podem ser utilizadas imagens, sons, símbolos ou outros elementos que representem sua necessidade

Esta pesquisa utiliza as formas textuais de representação, visto que são as mais utilizadas. Xie e Joo (2010) apresentam o uso de termos como a principal forma de interação no processo de busca de informações na Web, ou seja, a maioria das ações elaboradas estão associadas à formulação ou reformulação da consulta (ampliando, mantendo, restringindo ou alterando o contexto pesquisado). Ao abordar a elaboração de termos de busca em sistemas de recuperação de informação, Branski (2004) apresentou a atuação de sinais modificadores e dos operadores booleanos que podem ser utilizados na elaboração de expressões de busca. Importante ressaltar que a combinação correta de termos e operadores permite uma especificação mais adequada da necessidade informacional.

Em relação à formulação de expressão de busca, um estudo realizado por Jansen, Spink e Saracevic (2000) apresenta dados relevantes. Foram verificadas 51.473 consultas formuladas por 18.112 usuários do buscador Excite e elencados dados importantes sobre essas buscas. Destaca-se aqui parte da pesquisa que considerou a maneira como as expressões de busca foram formuladas. Nessa situação, os autores constataram que as expressões em geral são reduzidas, envolvendo em média 2,21 termos. Conforme apresentado pelo **Erro! Fonte de referência não encontrada.**, por volta de 62% das consultas observadas e analisadas continham um ou dois termos de busca e menos de 4% apresentaram mais do que 6 termos. No que diz respeito à estrutura das consultas, observou-se que a utilização de operadores (AND, OR, () e AND NOT) foi de apenas 8% das consultas realizadas. Termos modificadores (+, -, “ ”) apareceram em, aproximadamente, 15% das consultas.

Figura 6: Número de termos utilizados nas pesquisas do Excite

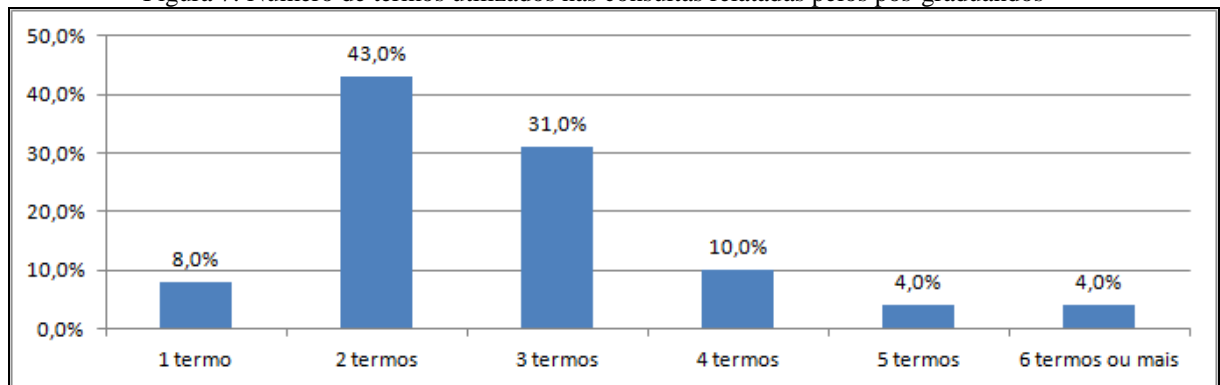


Fonte: Elaborado pelo autor baseado em Jansen, Spink e Saracevic (2000)

Segundo os autores, estes dados evidenciam que a maioria dos usuários realiza consultas mais simples, visto que uma porcentagem pequena de usuários utiliza operadores booleanos ou termos modificadores para elaborar suas buscas, além de utilizar poucos termos. Uma utilização coerente desses recursos possibilitaria a apresentação de melhores resultados de busca aos usuários.

Um estudo realizado por Giordano e Biolchini (2012) abordou o comportamento informacional de estudantes de pós-graduação em processo de recuperação na web. O estudo teve a participação de 52 estudantes de mestrado e doutorado. Foi identificado que o Google e o Google Acadêmico são as ferramentas mais utilizadas, sendo que 55,8 % dos mestrandos e 33,3 % dos doutorandos indicaram esta opção pela simplicidade e praticidade das ferramentas. Especificamente em relação à formulação das buscas, observou-se uma utilização mais significativa de termos modificadores, ou seja, cerca de 30% do público aplicou o operador "+" e sua variação "AND". O recurso de aspas duplas foi aplicado por 23,1%, com o intuito de manter uma ordem exata das palavras da expressão de busca. Conforme apresentado pelo **Erro! Fonte de referência não encontrada.**, a maioria dos participantes (43%) utilizou dois termos, sendo que a média de palavras utilizadas foi de 2,7. Cerca de 18% utilizaram mais de quatro termos na formulação de suas expressões.

Figura 7: Número de termos utilizados nas consultas relacionadas pelos pós-graduandos



Fonte: : Elaborado pelo autor baseado em Giordano e Biolchini (2012, p. 137)

Traçando um comparativo dos trabalhos de Jansen, Spink e Saracevic (2000) e Giordano e Biolchini (2012), observa-se algumas semelhanças e diferenças em relação à formulação de busca. Ambas utilizam linguagem textual e baseiam-se em termos de busca apresentando a média de pouco mais de 2 termos na formulação das expressões. Quanto à utilização de modificadores, houve significativa diferença. No primeiro estudo foi verificado que 15% utilizaram expressões contendo os caracteres (+, -, “ ”) e, no segundo, mais de 50% utilizaram tais recursos. Importante salientar que há uma diferença de público (público geral e pós-graduandos) e também de período (2000 e 2012), entretanto são dados importantes sobre como os usuários formulam suas expressões de busca.

Após a formulação da expressão de busca pelo usuário, é necessário que ela seja traduzida e representada de forma similar à utilizada na representação dos documentos. Essa semelhança visa possibilitar a comparação entre a busca e os documentos do *corpus* por meio da função de busca. Os modelos de RI a serem apresentados a seguir têm a função de realizar essa comparação e delimitam a forma de representação dos documentos e de representação da busca, neste caso, utilizando linguagem textual.

2.2.3 Modelos de Recuperação de Informação

A avaliação de relevância em um sistema de recuperação de informação é uma das questões centrais na tarefa de identificar documentos relevantes e descartar os irrelevantes. Para isto, os sistemas automatizados utilizam algoritmos com procedimentos e características previamente definidas para julgar a relevância e indicar uma ordenação utilizando critérios estabelecidos (BAEZA-YATES; RIBEIRO-NETO, 2013). A partir disso,

surgiram modelos de recuperação. Souza (2006) divide os modelos em clássicos e estruturados. Os modelos clássicos têm como característica básica a utilização de termos de indexação que procuram representar os documentos de forma sintética e significativa. Já os modelos estruturados utilizam, além dos termos, algumas informações sobre a estrutura do texto (como seções a serem pesquisadas, fontes de letras, proximidade das palavras, entre outras informações.).

São três os modelos clássicos de recuperação: o modelo booleano, o modelo vetorial e o modelo probabilístico. Esses modelos serviram como base para outros que visam estendê-los em funcionalidade e desempenho. A seguir serão descritos os modelos e seus fundamentos, os quais proporcionam características objetivas e palpáveis que serão utilizadas como suporte às ferramentas utilizadas e desenvolvidas ao final desta pesquisa. Esses modelos foram concebidos nos anos 1960 e 1970 e, posteriormente, aprimorados. As suas principais contribuições ainda estão presentes nas ferramentas de busca de informação.

2.2.3.1 Modelo booleano

O modelo booleano teve como base a teoria dos conjuntos e a álgebra booleana e é apresentado por Baeza-Yates e Ribeiro-Neto (2013) como um modelo simples, intuitivo e com semântica bastante precisa. Neste modelo os documentos são representados pelos seus termos de indexação. A expressão de busca pode ser composta por várias palavras ligadas pelos conectivos lógicos *AND*, *OR* ou *NOT*, formando uma expressão booleana que aplicada no conjunto de documentos, elenca quais satisfazem às restrições discriminadas.

Os conectivos *AND*, *OR* e *NOT* especificam respectivamente as operações de conjunção, disjunção e negação. Em sua pesquisa sobre recuperação de informação, Ferneda (2003) exemplificou, de forma simples e ilustrativa, a aplicação dessas conexões. O autor explica que uma expressão conjuntiva "t1 AND t2" recuperará documentos indexados por ambos os termos (t1 e t2). Já em uma expressão de disjunção "t1 OR t2" recuperará o conjunto dos documentos indexados pelo termo t1 ou pelo termo t2. E, por fim, a utilização da expressão "NOT t1" recuperará os documentos que não são indexados pelo termo t1. Caso a expressão não utilize os operadores e contenha apenas um termo "t1", terá como resultado o conjunto de documentos indexados por esse termo. A utilização de expressões mais complexas com diversos termos e operadores demanda um conhecimento aprofundado da lógica booleana, dessa forma seria possível realizar buscas mais detalhadas e restritivas.

Souza (2006) salienta que a maior desvantagem deste modelo é utilizar um

critério binário, tratando os documentos como relevantes ou não relevantes, não existindo formas de ordenação dos resultados. Dessa forma, o resultado da busca gera dois subconjuntos de documento. O primeiro são os que atendem ao critério especificado na expressão e o segundo os documentos que não se enquadram nos critérios. Mesmo este modelo apresentando limitações, Ferneda (2003, p. 51) esclarece que "muitos sistemas se desenvolveram utilizando o modelo booleano como ponto de partida para a implementação de novos recursos de recuperação".

2.2.3.2 Modelo vetorial

Baeza-Yates e Ribeiro-Neto (2013) relatam que o modelo vetorial apresenta uma boa técnica para tratar coleções genéricas de documentos. Por ser simples e rápido, é um modelo bastante popular no que diz respeito a características de ranqueamento. De acordo com Ferneda (2003), o modelo vetorial possibilita elencar documentos que correspondam, em parte, à expressão de busca. Os documentos são representados por um vetor numérico onde cada elemento representa o peso ou relevância do respectivo termo de indexação na representação do conteúdo informacional. Os autores destacam que os pesos do modelo vetorial são calculados através da ponderação TF-IDF, onde TF é a frequência do termo (*Term Frequency*) e IDF a frequência inversa do termo (*Inverse Document Frequency*).

Ferneda (2003) explica o TF como a quantidade de ocorrências de um termo em um documento. Ressalta-se que essa medida não considera o termo em questão em relação aos outros documentos da coleção. Acredita-se que se determinado termo ocorre em todos os documentos, ele terá pouco poder de identificação de relevância em relação aos outros. Conseqüentemente, é necessário determinar uma estatística global para a distinção de determinado termo em relação à coleção total. A medida denominada IDF tem objetivo de apresentar a ocorrência do termo analisado em relação aos documentos do *corpus*, ou seja, quanto menor o número de documentos contendo o termo, maior capacidade de evidenciar o documento que ele representa. Em resumo, "os melhores termos de indexação (os que apresentarão maior peso) são aqueles que ocorrem com uma grande frequência em poucos documentos" (FERNEDA, 2003, p. 33).

Da mesma forma que a representação dos documentos, as expressões de busca do usuário também são determinadas através de vetores numéricos multidimensionais. Dessa forma, a similaridade é calculada, matematicamente, através do produto escalar entre esses vetores. Os pesos de cada termo na representação também estão ponderados nos vetores

e após os cálculos de distância vetorial, ordenam-se os resultados apresentando os documentos com maior similaridade no topo da lista. O vetor de termos e seus respectivos pesos atribuídos na representação dos documentos e também na representação da busca são elementos fundamentais no desenvolvimento da proposta de interface desta pesquisa. Esses elementos atuarão como pontos de acesso, além de possibilitarem a percepção inicial do conteúdo dos documentos.

2.2.3.3 Modelo Probabilístico

Inicialmente o modelo probabilístico foi proposto em 1960 por Maron e Kuhns e, posteriormente, aproveitado por outros pesquisadores como S.E. Robertson e K. Sparck Jones, em 1976. Utiliza-se de conceitos da probabilidade para avaliar o grau de relevância entre a expressão de busca e o *corpus* documental. Correia (2016) esclarece que, de forma geral, este modelo procura estabelecer uma probabilidade de um documento ser relevante para uma formulação de pesquisa. Admite-se, neste caso, que a relevância depende da relação entre o documento e a requisição realizada pelo usuário. A ordenação é realizada por ordem decrescente da probabilidade de relevância.

Um dos fatores importantes desse modelo diz respeito à aplicação do conceito de *relevance feedback* que consiste em possibilitar aos usuários julgar relevância nos itens recuperados, ou seja, a partir da lista de documentos recuperados, o usuário seleciona os que considera relevantes e os submete ao sistema, o que possibilita listar resultados mais precisos. Pottker (2017) relata que por meio dessas interações os usuários podem alcançar, de forma gradativa, os resultados mais relevantes para sua necessidade de informação.

Os modelos clássicos são bastante objetivos e apresentam clareza nos seus critérios de representação documental e listagem dos documentos resultantes das buscas, portanto são utilizados para fundamentar diversos mecanismos de busca. Concluindo, o processo de recuperação de informação é suficientemente embasado pelos modelos clássicos. Ferneda (2013) explica que esses modelos utilizam os termos extraídos dos documentos e das buscas para efetivar o processo. O autor considera que este paradigma proporciona "um sistema fechado no qual o significado dos elementos lexicais é dado pelas suas inter-relações no interior de um corpus documental" (FERNEDA, 2013, p. 53).

2.2.4 Aspectos de Relevância na Recuperação de Informação

A ideia de relevância em um Sistema de Recuperação de Informação parte da relação entre as necessidades de informação dos usuários e os documentos que satisfaçam tais necessidades. Rijsbergen (1979) apresenta a relevância como um conceito subjetivo; dessa forma pressupõe-se ser extremamente complexa a sua aplicação nos sistemas de RI. Segundo Pavão (2014), fornecer ao usuário todos os documentos relevantes e ao mesmo tempo rejeitar todos os documentos irrelevantes tem sido um desafio para os motores de busca e os sistemas de recuperação de informação em geral. O autor descreve que os usuários que utilizam motores de buscas e sistemas de recuperação em bibliotecas, estão cada vez mais acostumados a formular consultas utilizando uma ou várias palavras em linguagem natural ou mesmo frases inteiras, dificultando o processo de RI.

Na opinião de Case (2012), o principal problema que enfrenta a operacionalização objetiva de relevância é a natureza contextual da opinião humana. Mesmo quando os documentos são ordenados a partir da terminologia utilizada na busca, pode haver razões para que não haja uma correspondência de julgamento sobre a verdadeira relevância dos documentos. Além do problema óbvio de que palavras podem ter significados múltiplos e variadas interpretações, tem sido efetivamente demonstrado que a avaliação de relevância pode mudar, dependendo da ordem em que são apresentados os documentos recuperados.

Em seus estudos, Pitangueira (2004) afirmou que a pesquisa de relevância seguiu uma linha geral que passa por tópicos como: o entendimento dos tipos de relevância (orientada ao sistema e orientada ao usuário), sua natureza, critérios, regiões e medidas de relevância. Esses aspectos estão resumidos no Quadro 2.

Quadro 2: Aspectos de relevância

Aspecto	Descrição
Tipos de Relevância	Conceito de relevância pode ser categorizado como orientado ao usuário ou orientado ao sistema
Crítérios de Relevância	A definição do conjunto de características dos objetos informacionais e de atributos da informação abriram perspectivas de pesquisa para a identificação dos critérios que os usuários utilizam no processo de RI.
Regiões de Relevância	É um domínio que serve de referência para avaliação das características dos objetos informacionais. Os usuários julgam os objetos em uma região de relevância sendo, na verdade, uma classificação multidimensional onde existem também os espaços parciais de relevância
Medidas de Relevância	Não há um consenso sobre o melhor sistema para a mensuração de julgamentos de relevância. Exemplos: (1-não relevante e 2-relevante) foram e são amplamente utilizadas; (1-não relevante, 2-relevância moderada e 3-relevante), entre outras gradações.

Fonte: Elaborado pelo autor baseado em Pitangueira (2004)

Ainda que o julgamento de relevância de um *corpus* documental seja considerado como uma das funções do usuário durante o processo de recuperação da informação, a medida de relevância também pode ser realizada por um sistema que qualifica se o documento é ou não pertinente à uma requisição de informação. Assim, esta medida deve basear-se na relação entre uma busca e o registro de documentos contidos numa coleção.

Saracevic (2007, p. 2131) analisou pesquisas sobre relevância comportamental, destacando alguns pontos:

- a) Individualmente as pessoas diferem em inferências de relevância, assim como eles diferem em todos os outros processos cognitivos em geral e que envolve informação, em particular;
- b) A relevância não é binária, ou seja, relevantes ou não relevantes. Os estudos apresentaram a necessidade de graduar documentos de acordo com sua relevância, isto é, os usuários elencaram os documentos em escalas de maior ou menor relevância;
- c) Relevância de objetos de informação não devem ser avaliados independentes, a relação entre os documentos associados deve ser considerada.

Os aspectos sobre relevância elencados nesta seção apresentam-se essenciais para o entendimento e contribuição no processo de RI, como Pitangueira (2004) salienta: "tanto o processo de julgamento de relevância como o comportamento do usuário têm profundo impacto no projeto de sistemas de recuperação de informação". Liu (2009) salienta que existem vários modelos para calcular a relevância entre consulta e um documento; alguns levam em consideração a proximidade dos termos da consulta e outros consideram a relação entre os documentos em termos de similaridade de conteúdo como estrutura de hyperlink, estrutura do site e diversidade de tópicos.

Nesta tese será considerado o modelo vetorial, apresentado no item 2.2.3.2 e que utiliza a similaridade dos documentos com a consulta realizada pelo usuário através de cálculos matemáticos entre vetores. A preferência por este modelo se explica por estabelecer um modelo conceitual que, ao contrário do modelo booleano, possibilita atribuir graus de similaridade que atuam como indicadores de relevância, permitindo aos resultados da busca serem ordenados. Também se verificou uma menor complexidade de implementação e processamento computacional em relação ao modelo probabilístico e a existência de soluções tecnológicas que implementam o modelo vetorial em seu núcleo.

Após apresentada a estrutura do processo de RI, os modelos e alguns

aspectos das concepções de relevância, torna-se necessário relacionar algumas publicações que descreveram desafios a serem enfrentados na recuperação de informação, em especial nas interfaces gráficas de visualização de resultados de busca. A escolha por alinhar a tese a essas interfaces proporciona um trajeto onde podem ser agregadas diversas vertentes de estudo que serão apresentadas oportunamente.

2.3 DESAFIOS DA RECUPERAÇÃO DE INFORMAÇÃO

A partir da apresentação da RI e da fundamentação inerente ao tema, foi possível elencar os estudos que exploraram a área no que diz respeito às interfaces de resultados de busca. Embora essas interfaces tenham evoluído, alguns problemas ainda perduram nos SRIs. Assim, julgou-se importante elencar esses problemas através de trabalhos que investigaram o tema.

Souza e Alvarenga (2004) exploraram brevemente os conceitos característicos aos sistemas de recuperação de informações no âmbito da Web Semântica. Os autores assim se colocam sobre o assunto: "A dificuldade de determinar os contextos informacionais tem como consequência a impossibilidade de se identificar de forma precisa a atinência dos documentos" (SOUZA;ALVARENGA, 2004, p. 133).

Em outro estudo sobre os sistemas de recuperação de informação e mecanismos de busca na *web*, Souza (2006, p. 163–164) também apresentou alguns obstáculos:

[..] a recuperação de informações traz dificuldades intrínsecas ao conceito de informação [..]. A associação entre os registros e seus conteúdos informativos é vaga, e isso pode acarretar problemas nas respostas a questões específicas, como baixas taxas de revocação¹ e precisão². Um sistema de recuperação de informações deve buscar boa relação entre os índices de revocação e precisão, para oferecer, em resposta a determinada consulta, referências ao maior número possível de documentos relevantes, ordenados por critérios que meçam esta relevância, e o menor número possível de documentos pouco ou não relevantes, de acordo com as necessidades de informação dos usuários.

Dias e Carvalho (2007) destacaram a sobrecarga de informações como uma das principais preocupações na representação dos resultados obtidos por meio de mecanismos de busca e recuperação de informações. Os autores explicam que atualmente os usuários

¹ A Revocação, ou "recall" ou mesmo "abrangeência", é a razão do número de documentos atinentes recuperados sobre o total de documentos atinentes disponíveis na base de dados. A revocação mede o sucesso do SRI em recuperar documentos pertinentes.

² Razão do número de documentos atinentes (relevantes) recuperados sobre o total de documentos recuperados.

acessam um grande volume de informações, muitas irrelevantes e desnecessárias ao seu contexto de interesse. Dessa forma, sugeriram utilizar formas gráficas de apresentação de informação baseadas na área de visualização de informações para minimizar esse tipo de problema.

A sobrecarga de informações também é citada por Case (2012). O autor explica que geralmente num resultado superior a 10 elementos, existe a propensão a essa sobrecarga, dessa forma é habitual que os usuários se satisfaçam com os primeiros resultados. Mesmo que não atendam plenamente às suas necessidades, esses resultados são aceitos simplesmente para abreviar o tempo de pesquisa. Essas informações são confirmadas por Jansen, Spink e Saracevic (2000), em sua análise do Excite, um serviço de busca na internet que utilizava o formato tradicional baseado em termos formulados pelos usuários. Os autores verificaram que os usuários têm tolerância limitada ao investigar lista de documentos; em geral são visualizados apenas os primeiros resultados. O dados coletados apontaram que, em média, os usuários examinaram e navegaram por 2,35 páginas de resultados, sendo que a maioria (58%) verificou apenas a primeira página.

Singh, Hsu e Moon (2013) argumentam que a estratégia mais comum na recuperação de informação tem sido a de apresentar os resultados como uma lista, onde cada entrada é classificada por sua relevância presumível para determinada consulta. Neste esquema, os usuários têm que posteriormente examinar a lista para satisfazer suas necessidades de informação. Os autores afirmam que tal processo é adequado para alguns tipos de consultas, no entanto, como a informação na web fica diversificada, este modelo de interação é, muitas vezes, insuficiente para descrever as informações que o usuário precisa, bem como os conteúdos que possam, supostamente, satisfazê-los.

Vieira e Correa (2011) descreveram técnicas propostas pela área de Visualização da Informação aplicadas na construção de interfaces gráficas, com o objetivo de proporcionar melhor apropriação da informação pelos usuários dos SRIs. No artigo destacaram três situações como as principais dificuldades na recuperação de informação:

- a) a sobrecarga de informações ou excesso de documentos retornados;
- b) a dificuldade do usuário em expressar ao sistema sua real necessidade por meio de palavras-chave;
- c) e a dificuldade do sistema em compreender a necessidade expressa e recuperar para o usuário os documentos que satisfaçam essa necessidade.

Na maioria dos mecanismos atuais de busca estão presentes alguns

problemas similares aos citados nesta seção. O Quadro 3 sintetiza alguns elementos importantes a serem explorados em relação às interfaces de resultados de busca em SRIs.

As obras apresentadas foram publicadas em revistas da área da Ciência da Informação, com classificação entre A1, B1 e B2 no Qualis³ da Capes (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior). Além da RI são abordadas temáticas como a Web Semântica e visualização da informação no âmbito da Ciência da Informação. Conforme já descrito, os problemas destacados no Quadro 3 servem como sustentação na busca por técnicas que possibilitem minimizar essas barreiras.

Quadro 3: Trabalhos que abordaram problemas na RI

Autores	Título	Problema destacado	Periódico /Qualis
Souza e Alvarenga (2004)	A web semântica e suas contribuições para a ciência da informação	Dificuldade de determinar os contextos informacionais.	Ciência da Informação (B1)
Souza (2006)	Sistemas de recuperação de informações e mecanismos de busca na web : panorama atual e tendências	A associação entre os registros e seus conteúdos informativos é vaga.	Perspectivas em Ciência da Informação (A1)
Dias e Carvalho (2007)	A visualização da informação e a sua contribuição para a ciência da informação	Sobrecarga de informações.	DataGramaZero Revista de Ciência da Informação (B1)
Vieira e Correa (2011)	Visualização da informação na construção de interfaces amigáveis para sistemas de recuperação de informação	Excesso de documentos retornados.	Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação (B1)
Singh, Hsu e Moon (2013)	Multiple perspective interactive search: a paradigm for exploratory search and information retrieval on the web	Modelo em lista é, muitas vezes, insuficiente para descrever as informações que o usuário precisa	Multimedia Tools Applications (B2)

Fonte: Elaborado pelo autor

2.4 SISTEMATIZAÇÃO E DISCUSSÃO

Estudar o processo de recuperação de informação mostrou-se uma abordagem que fornece princípios fundamentais para o desenvolvimento desta pesquisa. Neste sentido, apresentou-se a RI como uma área de pesquisa multidisciplinar, em especial

³ Qualis "é o conjunto de procedimentos utilizados pela Capes para estratificação da qualidade da produção intelectual dos programas de pós-graduação". A qualidade é determinada através de indicativos: A1 é o mais elevado; A2; B1; B2; B3; B4; B5; C - com peso zero. (CAPES, 2015)

das ciências da informação e da computação. Foram discutidos os aspectos históricos que proporcionaram condições de reconhecer a dimensão dos princípios elementares da área.

O desenvolvimento do capítulo alinhou-se ao processo de RI ilustrado por Ferneda (2013) e, nesta linha, foram analisados os conceitos relacionados aos elementos dispostos no processo. O processo considera dois extremos, ou seja, de um lado o acervo documental e, do outro, os usuários na busca por informação. Foram apresentadas as concepções de documento e suas formas de representação, em especial a indexação automática. Outro assunto descrito foi a conceituação de usuários em relação às suas necessidades informacionais, além de destacar o comportamento informacional e os modelos de comportamento de busca que sustentam a área.

Para que haja a interação entre os usuários e o acervo documental, os modelos de recuperação de informação fornecem subsídios com a finalidade de proporcionar essa comunicação. Foram destacados os modelos booleano, vetorial e probabilístico por fornecerem elementos simples e objetivos. Abordou-se, em seguida, algumas características do conceito de relevância que, embora complexo, é essencial. Por fim, baseando-se em pesquisas voltadas aos sistemas de recuperação de informação, foram descritas possíveis dificuldades e limitações desses sistemas. Concluindo, foi possível constatar a existência de diversos conceitos basilares na pesquisa por recursos que apoiem o progresso no processo de busca informacional.

3 INTERAÇÃO HUMANO COMPUTADOR - IHC

Na medida em que se busca tratar de questões envolvendo interfaces computacionais, surge a necessidade de analisar os conceitos relacionados à área de Interação Humano Computador (IHC), a qual abrange o projeto de interface e também os aspectos voltados à interação entre usuários e sistemas (PREECE et al., 1994). Neste sentido, torna-se coerente verificar métodos de interação e outros fatores da IHC com o intuito de analisar a viabilidade técnica da proposta. Este capítulo inicia-se com um breve histórico apontando fatos e autores importantes na delimitação da área; em seguida destacam-se alguns conceitos básicos (interfaces e interação) e são apresentadas as engenharias cognitiva e semiótica como bases teóricas da área. São descritos, também, aspectos relacionados às interfaces de recuperação de informação e os aportes fornecidos pela área de Visualização de Informação para o desenvolvimento de interfaces digitais.

3.1 FATOS HISTÓRICOS

O termo interação humano computador (IHC) foi cunhado na década de 1980 e delimitou uma nova área de estudo. Entretanto, diversos eventos anteriores contribuíram para a configuração da área, começando pela própria história da computação que tem como essência diversas referências, desde as primeiras representações numéricas, as primeiras tentativas de contagem e cálculo e as formas de automatização desses processos. Em relação às tecnologias de interação, Bush (1945) publicou um texto intitulado "*As we may think*", no qual apresentou diversas conjecturas para o futuro da ciência e em especial sua idealização de um dispositivo denominado "Memex". Neste dispositivo seriam armazenados livros, arquivos e comunicações que poderiam ser indexados e consultados de forma automatizada. Além disso, o autor descreveu como ponto forte de sua ideia o processo de relacionar dois ou mais elementos diferentes, realizando uma indexação associativa entre os conteúdos armazenados.

O dispositivo idealizado por Bush apresentou-se como referência para diversos autores que embasaram a criação das tecnologias de interação atuais no que diz respeito a dispositivos (*hardwares*) e programas (*softwares*). Ressaltam-se três personagens fundamentais. Conforme Cruz (2011), no início dos anos 1960, Douglas Engelbart foi um dos pioneiros da computação pessoal. Com base no Memex, o autor criou o mouse e colaborou na elaboração de uma rede de computadores pioneira na internet denominada ARPANET. Já

Theodore Nelson cunhou, em 1965, o termo hipertexto. Utilizando essa ideia, os leitores não seguiriam apenas um texto sequencial, podendo seguir conexões (*links*) citados para acessar outros conteúdos relacionados ao texto original. Por fim, em 1989, Tim Berners-Lee desenvolveu a linguagem HTML (*hyper text mark up language*) e os hyperlinks utilizados, atualmente, na Web. Esta tecnologia possibilitou a operacionalização das concepções iniciais do Memex de Bush, potencializando os elementos projetados por Nelson e Engelbart.

Prates e Barbosa (2007) destacam que durante a década de 1970 foram criados os primeiros computadores pessoais. Com o avanço no desenvolvimento das tecnologias de hardware, também começaram os investimentos em pesquisas para o desenvolvimento de sistemas úteis e voltados aos usuários. Seguindo essa rota, em meados dos anos 1980, a área de IHC se estabeleceu como área de estudo.

3.2 CONCEITOS BÁSICOS DE IHC

Preece et al (1994) explicam que a IHC preocupa-se com todos os elementos referentes à interação entre pessoas e computadores, sendo considerada uma área multidisciplinar contemplada pela ciência da computação, ciência da informação e ciências sociais e humanas. Considerando essa multidisciplinaridade da IHC, a definição de alguns conceitos básicos é inevitável para dar suporte ao desenvolvimento deste estudo. Assim serão explorados, a seguir, a aceção dos termos interface e interação, visto que são elementos basilares para a área.

3.2.1 Interfaces

A concepção de interface é usualmente utilizada para descrever o componente que interliga dois sistemas. Para Dias (1994), a interface é um modo de representar o modelo organizacional da informação; considera-se uma maneira de visualização do conteúdo e o ambiente que propicia o acesso a esse mesmo conteúdo. Santaella traz uma definição apropriada quando argumenta que "Interfaces são as zonas fronteiriças sensíveis de negociação entre o humano e o maquínico, assim como o pivô de um novo conjunto emergente de relações homem-máquina" (SANTAELLA, 2003, p. 92). Verifica-se que na interação homem-máquina considera-se interface o conjunto de *software* e *hardware* utilizados para permitir e facilitar o processo de comunicação entre o usuário e a aplicação.

Dale e Lewis (2011) consideram, de maneira simplificada, que os componentes físicos que permitem a realização de atividades motoras e perceptivas são denominados *hardware* e como exemplos tradicionais estão o teclado, mouse, monitor, entre outros. Já o *software* é responsável por estabelecer os processos computacionais que controlam o *hardware*, além de apresentar os elementos virtuais com os quais o usuário pode interagir utilizando-se de símbolos e mensagens para apresentar as informações do sistema. O *software* também realiza a interpretação dos comandos dos usuários, realizados através das mais variadas formas de interação.

Em conjunto com a área de IHC, houve uma evolução no desenvolvimento tanto de *hardware* como *software* para interfaces. Nos anos 1950 as interfaces eram painéis de controles do hardware, nessa época geralmente os usuários eram engenheiros. Entre os anos 1960 e 1980 as surgiram os primeiros aplicativos que eram manipulados por linguagens de códigos, ou seja, os usuários precisavam memorizar comandos para ativarem as funções pretendidas. A partir dos anos 1980 surgiram as primeiras interfaces gráficas com aplicativos e a maior utilização do mouse para interagir com tais aplicações. Os anos 1990 foram marcados pelo surgimento das interfaces multimídia e interfaces para *web*. A partir dos anos 2000 as interfaces começaram a se fazer presentes em diversos dispositivos além dos computadores pessoais, o que alavancou a necessidade de se pensar em formas diferentes de interação, um exemplo é a utilização de telas sensíveis ao toque as quais necessitam tanto de soluções de *hardware* como de *software*. Considera-se então ser importante a análise do conceito de interação.

3.2.2 Interação

Barbosa e Silva (2011, p.20) analisaram definições para interação e puderam verificar que houve uma evolução dessas definições ao longo dos anos. Inicialmente referia-se às sequências de estímulos e respostas entre corpos físicos, entretanto surgiram as pesquisas que não consideravam apenas a operação, e sim um modelo de comunicação entre usuários e máquinas. Posteriormente observa-se o processo pelo qual o usuário formula uma intenção atuando sobre a interface, podendo perceber e interpretar as respostas para avaliar se o objetivo foi atingido. Também se enfatiza a utilização da interação como processo de comunicação entre pessoas, mediada por sistemas computacionais. Por fim os autores consideram a interação usuário-sistema "como sendo um processo de manipulação, comunicação, conversa, troca, influência, e assim por diante". Constata-se, então, que as

abordagens privilegiam diferentes formas do fenômeno de interação.

Destacam-se, a seguir, os estilos de interação com os quais os usuários interagem com os sistemas computacionais, a saber: linguagens de comando, menus, WIMP (*Windows, Icons, Menus and Pointers*), preenchimento de formulário, manipulação direta e linguagem natural.

Uma interação baseada em **linguagem de comando** permite ao usuário redigir instruções para o sistema através da inserção de comandos específicos (PREECE et al., 1994). Neste modelo, os comandos podem ser realizados através de teclas de função, combinação de teclas e também digitação de palavras, abreviações ou, até mesmo, a utilização de um único caractere. Assim, é necessário memorizar os comandos que normalmente são definidos com base no vocabulário dos usuários, e a gramática da linguagem de comandos procura refletir a forma como eles conceituam as operações. Shneiderman (1998) destaca as seguintes características: precisão, concisão, facilidade de escrita e leitura, completude, rapidez no aprendizado, simplicidade para reduzir erros e facilidade de retenção ao longo do tempo. O autor indica que existe a correspondência entre a realidade e a notação dos comandos tentando refletir a estrutura da tarefa realizada. Como desvantagem, Oliveira (2010) explica que utilizar uma interface de linha de comando exige grande tempo e esforço para aprendizagem, principalmente para não especialistas.

Outra forma de interação é a utilização de **menus**. Neste esquema o sistema é responsável por oferecer uma lista de opções para que o usuário selecione a que lhe interessa. Souza et al. (1999) explicam que com a utilização de menus não existe a necessidade de memorização dos comandos, deve-se apenas reconhecê-lo, desta forma, os itens devem ser autoexplicativos. Quando existem muitos itens a serem exibidos é necessário um agrupamento por categorias e a utilização de submenus; também é possível ordenar os itens de forma alfabética, cronológica, itens mais frequentes, mais recentes ou mais importantes. Os autores consideram que este modelo tem a desvantagem de ocupar boa parte do espaço da tela, mas isto pode ser minimizado através da limitação da área de exibição.

Interfaces no estilo **preenchimento de formulários** são usualmente utilizadas para introdução de dados em um sistema de informação. São apresentados elementos para serem preenchidos que podem ser caixas de texto para digitação, lista de opções para seleção entre outros componentes que possibilitam informar dados ao sistema. Barbosa e Silva (2011) destacam a importância de agrupar itens relacionados, ordenando-os de forma lógica, além de apresentar instruções claras para o preenchimento. De modo geral os formulários são de simples entendimento, sendo possível que o sistema verifique os dados

antes de submetidos ao sistema para evitar possíveis erros na entrada de dados através de técnicas como a utilização de dígitos verificadores e totalização de valores.

A sigla **WIMP** (*Windows, Icons, Menus and Pointers*) que pode ser traduzido por janelas, ícones, menus e apontadores, permite a interação por meio de componentes virtuais exibidos em interfaces gráficas. Os componentes e suas funções são manipuladas através dos dispositivos de hardware como o teclado, mouse ou a própria tela do dispositivo. Souza et al. (1999) descrevem que o WIMP é uma combinação de hardware e *software* associado aos outros estilos apresentados (menus, preenchimento de formulário, manipulação direta e linguagem de comandos) e proporcionam fácil utilização e aprendizado. Para Oliveira (2010) este modelo de interação é o mais utilizado pelos variados tipos de usuários e está presente em praticamente todos os dispositivos computacionais. O autor aponta como desvantagem da interface gráfica o grande número de distrações que há na tela.

Conforme discutido por Shneiderman (1998), a **manipulação direta** oferece aos usuários a possibilidade de agirem diretamente em objetos de interesse exibidos, usando ações como, por exemplo, arrastar e soltar. O objetivo é auxiliar usuários a utilizarem a interface tendo o mínimo de aprendizado. Este estilo de interação fornece *feedback* imediato para as ações físicas e suporta ações facilmente reversíveis que promovem a exploração no ambiente digital. Assim, os comandos baseiam-se na analogia entre o cursor e a mão e as representações gráficas de objetos dispostas na tela; em geral utiliza dispositivos como o mouse ou toque diretamente na tela do dispositivo, utilizando telas sensíveis ao toque. Para aplicativos de recuperação de informações, Marchionini (2005) ressalta que Shneiderman e seus colegas desenvolveram interfaces de consulta dinâmica que permitem que as pessoas façam consultas simplesmente movendo o mouse e observando, imediatamente, os resultados.

A **linguagem natural** é um estilo de interação que objetiva proporcionar ao usuário se expressar utilizando o próprio idioma. É uma forma bastante atrativa para usuários com pouco conhecimento em computação. Conforme apresentado por Barbosa e Silva (2011), este estilo visa facilitar a utilização de um sistema por usuários inexperientes, mas salientam alguns desafios para o desenvolvimento de interfaces capazes de estabelecer significados resolvendo ambiguidades e imprecisões nas proposições dos usuários. Os sistemas de RI buscam utilizar este tipo de interação em seus processos, visando proporcionar a interação através de linguagem natural. Para isto pode-se fornecer dispositivos que possibilitem a digitação de frases ou a gravação e reconhecimento da fala dos usuários com o intuito de aproximar a aplicação do usuário, privilegiando a forma de comunicação deste.

Concluindo, esta apresentação dos estilos de interação visou fornecer

subsídios para o entendimento das possíveis utilizações na proposta da interface aqui discutida.

3.3 BASES TEÓRICAS DA IHC

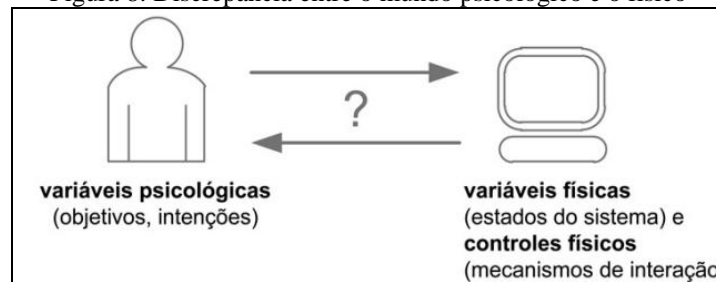
A IHC contém bases teóricas para fornecer explicações e previsões dos fenômenos na interação entre o usuário e as interfaces digitais. Conforme será explanado nesta seção, as Engenharias Cognitiva e a Semiótica se complementam para embasar o desenvolvimento de propostas de interação, verificando os aspectos cognitivos do usuário em relação às interfaces que são consideradas artefatos de metacomunicação pela semiótica.

3.3.1 A Engenharia Cognitiva

Em 1986, Norman (1986) propôs a Engenharia Cognitiva considerando-a como uma ciência cognitiva aplicada, objetivando utilizar os conceitos dos estudos cognitivos e aplicá-los no desenvolvimento de dispositivos eletrônicos. Na época, o autor diagnosticou certas dificuldades para o uso dos computadores e vislumbrou aplicar os conhecimentos da ciência cognitiva para aprimorar este panorama. Portanto, o objetivo da Engenharia Cognitiva é compreender os processos psicológicos envolvidos no uso de computadores, relacionando-os aos fenômenos envolvidos durante a interação com o sistema.

Para Barbosa e Silva (2011), na base da engenharia cognitiva existe uma discrepância (Figura 8), entre os objetivos psicológicos e as variáveis físicas de uma determinada operação. Por fatores psicológicos entende-se os objetivos e intenções dos usuários, de acordo com as necessidades informacionais. Como variáveis físicas denotam-se os controles a serem manipulados no sistema, os quais afetam o estado do sistema e proporcionam a operacionalização das tarefas. Assim os usuários necessitam traduzir as suas intenções em ações físicas sobre o sistema. A discrepância indicada por Barbosa e Silva (2011) diz respeito à necessidade do usuário entender os controles e variáveis físicas e traduzir em ações que o levem a atingir seus objetivos no sistema.

Figura 8: Discrepância entre o mundo psicológico e o físico



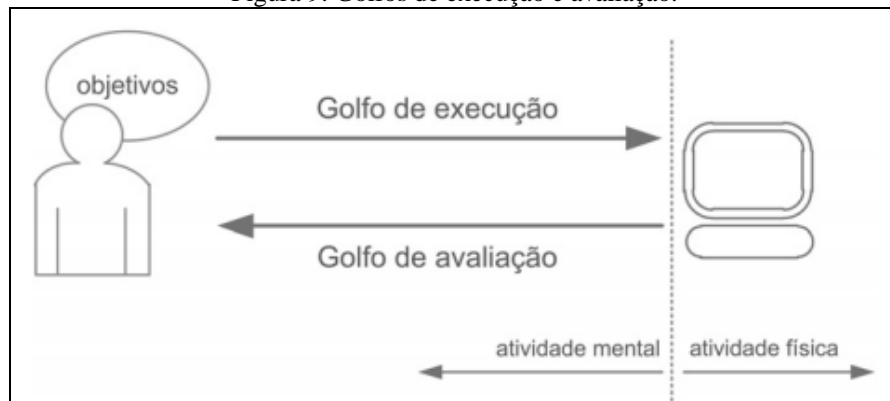
Fonte: (BARBOSA; SILVA, 2011, p. 53)

A Figura 8 sugerida por Barbosa e Silva (2011) ilustra que entre as variáveis psicológicas e as físicas deve haver um estágio de interpretação para relacioná-las. Também é necessário que existam funções que relacionem a manipulação dos elementos físicos e os resultados das ações. Seguindo a mesma linha, Norman (1986) explica que a Engenharia Cognitiva concentra-se na noção de que a IHC é pautada pela interpretação e avaliação de atividades executadas pelos usuários que possuem o estímulo de concretizar necessidades em eventos de entrada e interpretar as respostas do sistema. Entende-se importante compreender o processo de cognição humana para se explorar os aspectos cognitivos existentes nas interfaces computacionais. Os fatores cognitivos são executados através de métodos ou características na construção das interfaces, com o intuito de torná-la adequada ao uso, facilitando o processo de aprendizado do usuário.

A Engenharia Cognitiva pode ser representada através do modelo GOMS – *Goals, Operators, Methods and Selection rules* (Metas, Operadores, Métodos e regras de seleção). Este modelo baseia-se na ideia que os usuários têm metas a cumprir, e que para atingir essas metas escolhem dispositivos para operacionalizar as atividades e o método como irão realizar. No que diz respeito à discrepância descrita anteriormente, as metas são consideradas variáveis psicológicas e deverão ser atingidas utilizando-se dos mecanismos físicos.

A interação entre os indivíduos e o sistema delimita dois alvos a serem atingidos. Norman (1986) define esses dois alvos como "golfos" a serem superados, ou seja, o golfo de execução e o de avaliação. O golfo de execução é considerado a especificação da intenção e da série de ações realizadas na interface do sistema. Por sua vez, o golfo da avaliação é o confronto da saída do sistema (sua resposta) com a meta do usuário para aquela ação (a Figura 9 representa essa situação). Em resumo, o processo de interação com os sistemas pode ser considerado uma série de ciclos de ação envolvendo fases de execução e de avaliação, alternadamente.

Figura 9: Golfos de execução e avaliação.



Fonte: (BARBOSA; SILVA, 2011, p. 56)

Exemplificando o processo de forma simples, o usuário define uma meta, que poderia ser encontrar um artigo sobre Ciência da Informação. Em seguida, utiliza a interface física do sistema para realizar esta busca utilizando os componentes oferecidos a ele tanto de hardware (mouse, teclado, monitor, entre outros), como de *software* (caixas de texto, botões, barras de rolagem, etc.). Após executar a busca, recebe os resultados e os compara com sua meta. Caso a meta tenha sido atingida, o processo se encerra; caso contrário, realiza-se um novo ciclo até que sejam atingidos os objetivos. A Engenharia Cognitiva centra-se em verificar esses ciclos e tentar disponibilizar ferramentas que possibilitem ao usuário atingir seus objetivos de forma mais rápida e intuitiva.

A Engenharia Cognitiva observa a interação dos usuários e sistema, ou seja, o sistema é elaborado com foco em disponibilizar formas de executar comandos e funções em que os usuários tenham facilidade de utilização. Entretanto, Netto (2010) esclarece que o processo de desenvolvimento e projeto de interfaces não é estudado pela Engenharia Cognitiva e sim pela Engenharia Semiótica, que será apresentada a seguir.

3.3.2 A Engenharia Semiótica

Os primeiros estudos sobre a teoria da Engenharia Semiótica (ES) iniciaram-se na década de 1990, mas sua primeira versão foi publicada, internacionalmente, no ano de 2005, por Clarisse S. de Souza. Os fundamentos da ES estão pautados na Semiótica, uma área que analisa fenômenos de comunicação e significação, detalhados no Capítulo 4. De maneira geral a semiótica tem por objetivo "o exame dos modos de constituição de todo e qualquer fenômeno de produção de significação e sentido" (SANTAELLA, 2012a, p. 19). Por sua vez o termo "Engenharia" é conceituado por Scottini

(2009) como a arte de projetar, desenhar e construir. Assim a Engenharia Semiótica atua na IHC oferecendo "conceitos e métodos para apoiar o processo comunicativo desde sua elaboração, bem como para embasar uma análise da consistência e coerência deste processo" (LEITÃO, 2013, p. 5).

Conforme conceituado por Souza (2012), a ES parte do princípio de que a Interação Humano-Computador constitui-se de uma comunicação humana intermediada por computador. Salienta-se que os projetistas de interface participam ativamente do processo de comunicação utilizando a interface projetada como meio para troca de mensagens com o usuário. Para isso utilizam-se de signos como palavras, ícones, sons e controles na interface para apresentar sua visão aos usuários. Do outro lado os usuários interpretam tais signos no momento da interação com o sistema. Verifica-se que nessa visão os usuários se comunicam com o sistema em um processo de metacomunicação, a interface apresenta a visão do projetista sobre o que ela pode oferecer, quais problemas resolver e como realizar as interações, assim as mensagens trocadas entre usuário e interface possibilitam a execução das tarefas desejadas.

O conceito de metacomunicação considerado é descrito por Leitão, Silveira e Souza (2013, p. 12) como

[..] o discurso interativo do designer a respeito de quem é o usuário visado e como, quando e onde ele pode se comunicar com o sistema, para que finalidades e com que vantagens. Compreende o conjunto de todas (e somente) as conversas que o usuário pode ter com o sistema em tempo de interação. Em tempo de design, a mensagem de metacomunicação é antecipada e projetada integralmente, e em seguida especificada e codificada computacionalmente pelo desenvolvedor.

Souza (2012) ratifica essa ideia explicando que os projetistas não estão presentes no momento em que o usuário interage com o *software*, por isso necessitam estar representados na interface. Deve utilizar-se um sistema de significação que proporcione aos usuários saber o que o *software* faz, como e por quê. Existem várias formas de representações, ou seja, os projetistas podem se apresentar como humanoides (interfaces de sistemas com características humanas com habilidades de afeto e linguagem natural), como máquinas (interfaces de sistemas com botões de pressão, controles deslizantes, mostradores e similares), ou mesmo como espaços (sistemas com mundos virtuais que os usuários exploram para obter vários tipos de efeitos). Para a autora, de acordo com a mensagem, algumas representações atuarão melhor que outras. Como exemplo, possivelmente os humanoides transmitam explicações e instruções mais adequadamente do que os espaços virtuais. Presume-se também que as máquinas podem transmitir capacidades físicas com mais eficiência do que o discurso

da linguagem natural dos humanoides.

A definição desse esquema de representações a ser utilizado é um processo de engenharia de signos (ou engenharia semiótica). Para Leitão, Silveira e Souza (2013) podem ser utilizados signos conhecidos, novos signos ou associar novos significados a signos conhecidos. A Engenharia Semiótica distingue três classes de signos na organização e expressão do projetista na interface: signos estáticos, signos dinâmicos e signos metalinguísticos:

- a) signos estáticos são signos que expressam um estado do sistema e cujo significado pode ser entendido sem a necessidade de se interagir com o sistema. Como exemplos citam-se os elementos dispostos em uma interface como botões e caixas de texto.
- b) signos dinâmicos são signos que expressam o comportamento do sistema e cujo significado pode ser compreendido no decorrer da interação com o sistema. São exemplos: o deslocamento do foco da entrada de dados durante o preenchimento de um formulário e a ativação e desativação de um botão de comando.
- c) signos metalinguísticos são signos principalmente verbais e que se referem a outros signos de interface, sejam eles estáticos, dinâmicos ou mesmo metalinguísticos. São exemplos desse tipo de signo as mensagens de ajuda e de erro, alertas, diálogos de esclarecimento, dicas e assemelhados.

A exposição acima permite destacar o papel ativo do projetista no processo de interação. Na visão de Barbosa e Silva (2011), o projetista tem obrigação de ajudar os usuários a compreenderem a metamensagem, apresentando a sua visão sobre as necessidades do usuário e, principalmente, as questões de projeto para o entendimento das funções da interface. Para isso é primordial a escolha adequada de estratégias comunicativas, seleção dos tipos de signos a serem utilizados na linguagem de interface e as limitações que os significados computacionais trazem para a interação.

3.4 PRODUÇÕES CIENTÍFICAS SOBRE INTERFACES DE RECUPERAÇÃO DE INFORMAÇÃO

Após a apresentação da área de IHC, através dos conceitos iniciais e bases teóricas, busca-se proporcionar uma visão de pesquisas que apresentam interfaces de RI, em

especial interfaces de resultados, verificando características e possíveis implicações para este estudo. Buscou-se as produções científicas realizadas no ano de 2018 com o intuito de delimitar os estudos mais recentes nas áreas de interesse. Para isso utilizou-se como fonte de pesquisa o Portal de Periódicos da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Capes) que disponibiliza acesso a bases nacionais e internacionais de pesquisa científica.

Foram utilizados os termos "interfaces" e "recuperação de informação" além de suas variações na língua inglesa. Inicialmente foram verificados os resumos que proporcionaram filtrar as pesquisas que apresentaram relação com esta tese.

Inicia-se esta discussão através do estudo realizado por Sedghi, Zeinab e Tahamtan (2018) que analisaram a interação dos usuários com as interfaces de busca visual, abordando fatores que constituem o contexto deste tipo de interação. Foi realizada uma pesquisa qualitativa através de entrevista com 28 indivíduos (20 estudantes de mestrado e doutorado e 8 bibliotecários), analisando a interação dos usuários com a pesquisa visual da EBSCO (instituição fornecedora de bases de dados para pesquisa). Embora o sistema proporcione apresentar os resultados no formato textual, no estudo esses resultados de busca foram exibidos utilizando recursos como fontes claras e escuras, nas cores verde e azul e na forma de um fluxograma em uma página.

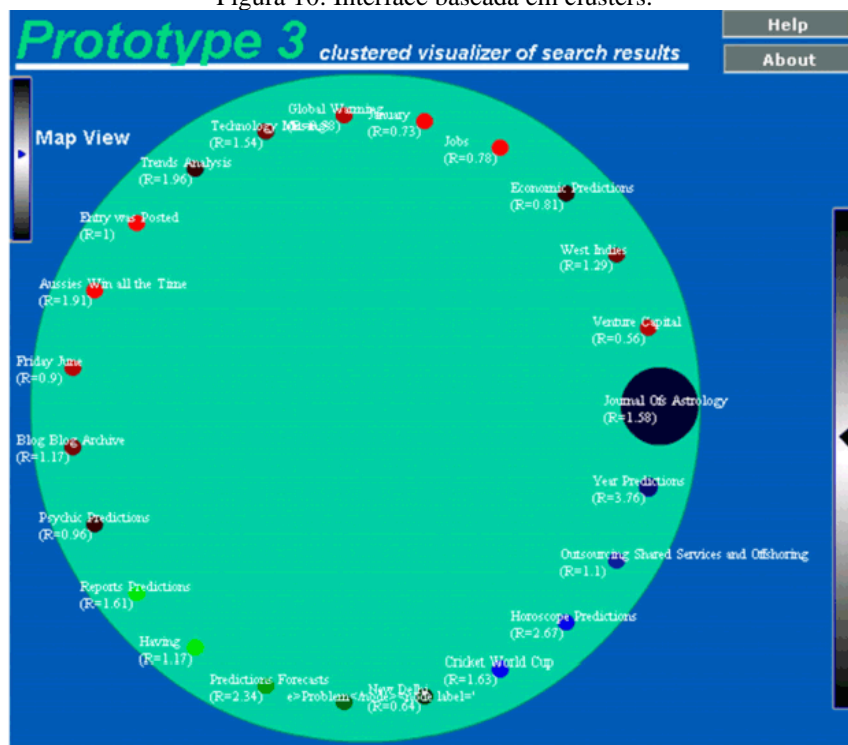
Como apontamentos da pesquisa, Sedghi, Zeinab e Tahamtan (2018) indicaram a preferência dos usuários por sistemas de pesquisa com os quais já estão familiarizados (baseados em palavras-chave), isto devido à simplicidade e de suas experiências nesse tipo de ferramenta. Entretanto os usuários que já utilizavam a ferramenta visual da EBSCO realizaram pesquisas mais eficientes, foram mais bem-sucedidos no atendimento de suas necessidades de informações e atestaram interações positivas com a interface de pesquisa visual. Também foi descrito que a utilização de imagem em formato de um quadrado para apresentar os artigos mais relevantes foi bem avaliada pelos usuários; já o tempo de resposta do sistema para apresentar os efeitos visuais e, por consequência, proporcionar acesso aos documentos foi apontado como um dos problemas da interface. Por fim, constatou-se que as interfaces de busca visual possuem vantagens e desvantagens e um dos fatores que afetam o uso é a experiência do usuário com a ferramenta. Os autores estimulam o desenvolvimento de outros estudos que possam investigar formas de interação do usuário com a interface de pesquisa visual para proporcionar facilidade e velocidade no acesso às informações.

Uma pesquisa sobre o impacto das inovações no design da interface de

usuário em mecanismo de busca foi realizada por Tomasi, Schuff e Turetken (2018). Os autores analisaram a eficiência da exibição gráfica de resultados. Como justificativa para a investigação apresenta-se o fato de que embora existam evidências que inovações na interface utilizando exibições gráficas possam produzir ganhos de eficiência no processo de busca, até o momento não se apresentaram alternativas comerciais bem-sucedidas. Dessa forma, foi conduzida uma experiência controlada em laboratório que apresentou uma nova interface para os usuários que navegam pelos resultados de pesquisa na web. Na experiência foram apresentados dois modelos de interface, uma mais tradicional no formato de lista de resultados como no Google; a segunda interface apresentava graficamente os resultados da pesquisa em clusters, conforme Figura 10.

No relato da pesquisa, Tomasi, Schuff e Turetken (2018) verificaram que os usuários com menos experiências em tarefas de buscas estavam mais propensos a experimentar as inovações; a explicação para isso vem da ideia de que eles não formaram seus esquemas e por isso podem estar mais abertos a novos recursos. Esse tipo de situação corrobora o entendimento de que os indivíduos adotam estruturas habituais para realizarem suas tarefas rotineiras e isso causa certo desconforto com a utilização de novas ferramentas. Indica-se, então, que novas formas de interface sejam implementadas de forma gradativa, atuando como formas alternativas de acesso à informação.

Figura 10: Interface baseada em clusters.



Fonte: (TOMASI ET AL., 2018, p.411)

Em pesquisa recente realizada por Hoeber (2018), foram destacados aspectos que envolvem a RI em conjunto com a área de visualização de informação. Para isso o autor considerou tanto as características dos dados quanto as necessidades dos usuários, verificando as codificações visuais que permitem aos usuários perceber, interpretar e dar sentido às informações apresentadas a eles. Foram elencadas as variáveis visuais que podem ser manipuladas e mapeadas para dimensões e aspectos específicos dos dados envolvendo posição, tamanho, forma, brilho, textura, matiz e orientação. O autor ressaltou que a escolha de quais aspectos pretende-se mapear para as variáveis visuais também depende do tipo de dado (quantitativo, ordinal, qualitativo) e da tarefa pretendida pelo usuário (associação, seleção, ordem, quantidade).

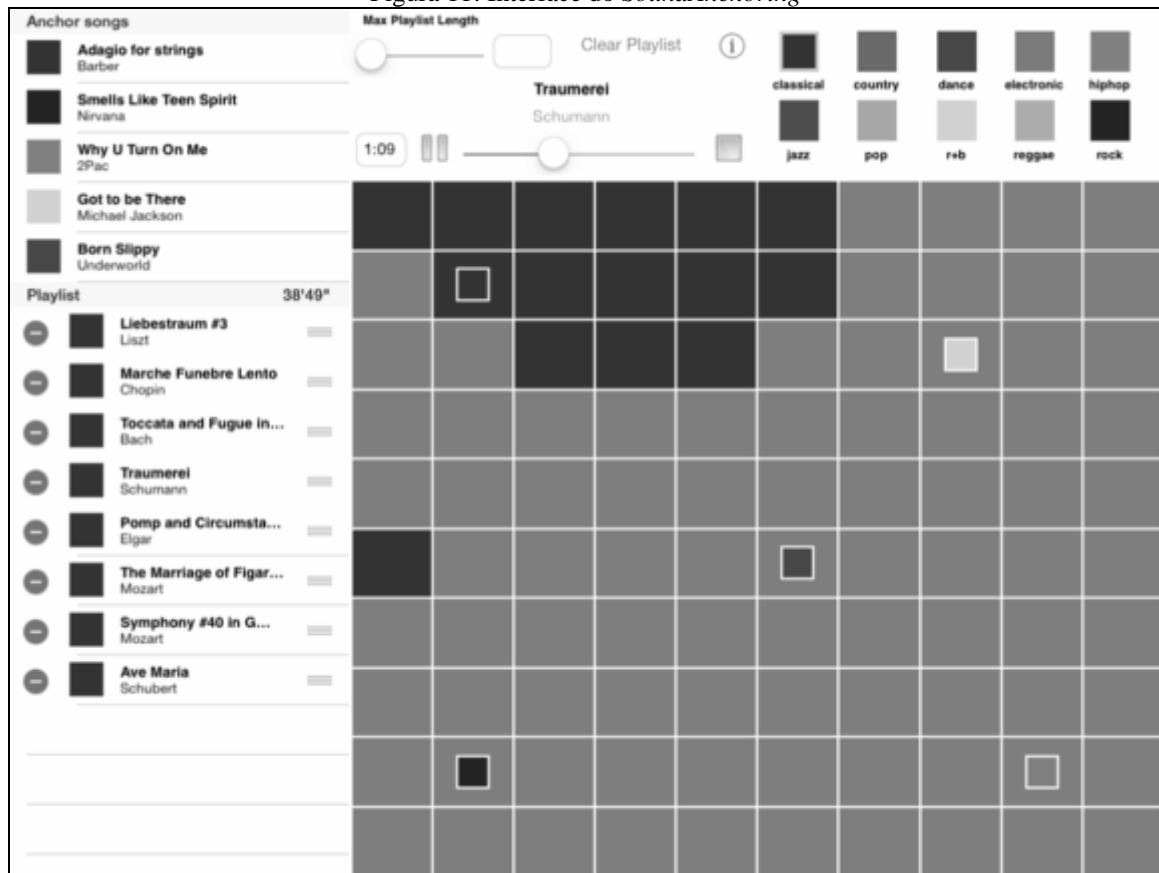
Para fundamentar suas contribuições, Hoeber (2018) se baseou em algumas áreas como teoria das cores, princípios da Gestalt e a teoria do forrageamento (*Information foraging theory*), adaptado para informação (forragear significa vasculhar, remexer, à procura de algo). No estudo indica-se que técnicas de visualização de informação podem ser usadas para mostrar as relações entre os resultados da pesquisa e os diferentes aspectos da consulta. Além disso, seria possível fornecer visões que permitam aos indivíduos avaliar rapidamente a composição geral dos resultados utilizando-se de artifícios para apresentar as relações entre os resultados da busca e os diferentes aspectos da consulta. Por fim, o autor fomenta a necessidade de pesquisas que envolvam o uso de técnicas de visualização de informações no desenvolvimento de novas interfaces de pesquisa e que estejam fundamentadas em teorias e princípios estabelecidos. Uma fundamentação adequada permite um estudo mais aprofundado sobre como a visualização da informação pode apoiar comportamentos e estratégias de busca de informação.

Os estudos na área de RI também contemplam ideias relacionadas à busca de informação em formatos de mídia. Embora esta pesquisa aborde basicamente o sistema de busca textual, outras ideias com propostas de interfaces são bastante úteis para verificar o panorama recente das pesquisas na RI. Neste sentido, Collares et al. (2018) propuseram um sistema denominado *SoundAnchoring* para visualização baseada em conteúdo e exploração de coleções de música com base no algoritmo *Self-Organizing Map* (SOM). O objetivo do sistema é permitir que os usuários escolham os locais dos clusters que contêm faixas semelhantes, acusticamente, no espaço musical.

O *software* desenvolvido por Collares et al. (2018) foi dividido em três etapas: extração de recursos, organização e visualização. Durante a extração de recursos as músicas são analisadas e são extraídas informações de trechos de cada música, dessa forma

esses dados são salvos em diversos vetores que são utilizados posteriormente para comparação de padrões e no processo de recuperação. Na etapa de organização o algoritmo SOM é empregado para mapear o espaço de recursos em uma representação bidimensional que pode ser exibida em uma tela. A topologia de dados é preservada, ou seja, vetores de características semelhantes devem ser colocados próximos uns dos outros. Por fim a etapa de visualização consiste em exibir a saída do algoritmo denominado "*Anchored SOM*" (adaptação do algoritmo SOM) e fornecer aos usuários ferramentas para explorar a coleção de músicas.

Para chegar à interface final (Figura 11) que possibilita acesso à coleção musical, os usuários navegam por uma sequência de telas e indicam escolhas que influenciam a organização e a aparência do espaço musical. O Nesta interface final são apresentadas pelo *software* "músicas âncoras" determinadas através das escolhas do usuário. As âncoras são listadas no canto superior esquerdo e suas localizações na grade são representadas como quadrados. A lista de reprodução é mostrada abaixo da lista de âncoras. A categorização de gênero visa permitir que os usuários interajam com subconjuntos da coleção de músicas. A grade exibe apenas trilhas clássicas (nós mais escuros), já que os outros gêneros foram desativados.

Figura 11: Interface do *SoundAnchoring*

Fonte: (COLLARES ET AL., 2018, p.5537)

Após a implementação foi realizado um estudo de usuário para avaliar a aplicação desenvolvida baseado em tarefas e cenários. Como resultado, esse estudo sugeriu que a ancoragem proposta no *software* não teve impacto na qualidade da lista de reprodução, mas a disposição das músicas âncoras possibilitou uma interação mais atrativa e fácil de controlar. Também se verificou que a criação de listas de reprodução se tornou mais rápida. Embora os autores apontem para a necessidade de realização de mais testes, concluiu-se que a exibição das músicas em uma interface não tradicional tem um grande potencial a ser explorado na tentativa de melhorar a interação dos usuários com grandes coleções de músicas.

Durante a busca por trabalhos relacionados, identificou-se um artigo formulado por Muller et al. (2018) que descreveu o desenvolvimento do "Textpresso" que consiste em um sistema automatizado de extração de informações na área biomédica. O objetivo é extrair com eficiência o texto completo de artigos de periódicos para informações biológicas. Ainda que o trabalho não apresente uma interface que utiliza elementos visuais, o *software* utiliza a tecnologia Lucene que serve como alicerce para produção de sistemas de RI.

Os autores relatam que o Lucene fornece o processo de indexação e busca

que permite a manipulação eficiente de grandes volumes documentais a partir de textos completos. O processo de indexação apontou para uma taxa de cerca de 30 artigos por minuto e núcleo do processador. Os artigos ficam disponíveis para a busca no módulo de pesquisa, onde a busca é feita baseada em palavras-chave e os resultados geralmente classificados por pontuação, que é calculada pelo Lucene através de algoritmos que relacionam a frequência de termos e a frequência inversa dos documentos (TF/IDF). Por fim, é interessante salientar que a utilização desta tecnologia proporciona a estrutura necessária para um processo de recuperação de informação eficiente que atende tanto a indexação quanto a busca apoiada em palavras-chave.

O objetivo desta seção foi apresentar trabalhos recentes sobre temas relacionados e que podem ocasionar alguma implicação para o desenvolvimento de interfaces em sistemas de recuperação de informação. Foi possível levantar vantagens e desvantagens, sendo a principal barreira a resistência dos usuários a mudanças em um processo a que já estão acostumados. Por outro lado, a ideia de viabilizar alternativas é interessante e deve ser realizada de forma gradativa, indicando que o uso de técnicas de visualização de informação são formas válidas para melhorar a experiência do usuário.

3.5 VISUALIZAÇÃO DE INFORMAÇÃO

A Visualização de Informação (VI) se associa à IHC no sentido de buscar formas de otimizar a recepção de informações por parte dos usuários. Dias e Carvalho (2007) expõem que a associação dessas áreas visa elevar a qualidade na construção de interfaces tendo em vista que a VI geralmente é um objeto inserido nas interfaces estudadas pela área de Interação Humano Computador. Nascimento e Ferreira (2005) explicam que o campo coberto pela VI estuda a construção de representações visuais para dados abstratos com o objetivo de facilitar a compreensão e esclarecimento de informações contidas nesses dados. A Visualização da Informação é conceituada por Freitas et al. (2001, p. 143) como

[..] uma área da ciência que tem por objetivo o estudo das principais formas de representações gráficas para apresentação de informações, a fim de contribuir para o entendimento delas, bem como ajudar a percepção dos usuários a fim de deduzir novos conhecimentos baseados no que está sendo apresentado. É uma ciência que combina aspectos de computação gráfica, interação humano-computador, cartografia e mineração de dados.

O campo da VI tem história recente. De acordo com Xavier (2009), a área é um notório campo de pesquisa e tem crescido rapidamente devido a sua extensibilidade em

campos de pesquisa interdisciplinares. O autor destaca que os estudos de VI convergem com trabalhos na IHC abrindo possibilidades de aplicação desses conceitos na recuperação de informação. Conforme afirmação de Santos e Gaspar (2013, p. 11), é primordial a utilização de novas formas de apresentação de informação como alternativa ao formato de lista textual, pois dessa forma a disseminação de informação se torna mais eficiente. Neste cenário, novos modelos de interfaces podem alterar o modo como é realizada a visualização e a interação com informações recuperadas. Em sua tese, Valiati (2008, p. 12) avaliou a usabilidade de técnicas de visualização de informações e argumenta:

[...] as técnicas de visualização de informações multidimensionais têm o potencial de auxiliar na análise visual e exploração de grandes conjuntos de dados, através do emprego de mecanismos que buscam tanto representar visualmente os dados quanto permitir ao usuário a interação com estas representações.

A autora explica que na apresentação dos documentos são criados modelos gráficos ou representações visuais dos dados que suportam a interação direta do usuário na exploração e aquisição de informações úteis neles contida. Ela também descreve que projetar interfaces com níveis aceitáveis de usabilidade é uma tarefa complexa, de grande amplitude, que envolve inúmeras questões, pois as necessidades e o público-alvo variam de aplicação para aplicação. Para tanto, várias soluções têm sido propostas, incluindo o uso de heurísticas no projeto, especificação com modelos e métodos formais, adoção de normas, guias de recomendação e técnicas de avaliação de usabilidade.

Na pesquisa sobre o contributo da Visualização de Informação para a Ciência da Informação, Dias e Carvalho (2007) realizaram alguns experimentos práticos para demonstrar a importância do uso de estruturas de visualização da informação na transmissão de conhecimento. Em suas conclusões afirmaram que as estruturas de Visualização da Informação podem contribuir para a tomada de decisão, descoberta de novos conhecimentos, demonstração de esquemas, representação de ideias e análise das informações que podem tornar mais ágil a apropriação de conhecimento por parte do usuário.

Vieira e Correa (2011) avaliaram requisitos da representação visual de informações na construção de interfaces amigáveis para SRIs. Na opinião dos autores, o desenvolvimento de interfaces embasadas na VI possibilita um melhor entendimento da organização da informação, possibilitando uma visão geral dos assuntos, a descoberta de novos conceitos e a contextualização dos resultados de busca através da relação entre esses resultados. Vaz e Carvalho (2004) destacam que a principal finalidade da área de Visualização de Informações é absorver informações, o máximo possível, de um grupo de

dados. Logo, esta representação é disponibilizada de diversas formas: gráficos, tabelas, grafos, entre outros. Deste modo, o usuário é capaz de adquirir um conhecimento às vezes implícito nos dados. Explorando-se então os diversos fatores que impactam na recuperação e visualização de informação, verifica-se a possibilidade da união destes conceitos, com o objetivo de proporcionar maior eficiência na RI.

O processo de busca por informação apresenta desafios e oportunidades para visualização na recuperação de informação. Usando técnicas de visualização de informações, as pessoas podem efetivamente visualizar as relações entre documentos e os termos que os representam. Para Song (2000) as técnicas de VI são utilizadas para apresentar medidas de similaridade e distância de documentos. Em geral os documentos são posicionados na forma de um gráfico ou mapa de dispersão, agrupando documentos automaticamente através de métodos estatísticos. Os mecanismos de pesquisa podem retornar as consultas realizadas pelos usuários com um grande número de documentos, tornando difícil identificarem os relevantes. Nesse contexto, as técnicas de visualização de informações fornecem possibilidades de resolver tais situações.

Algumas técnicas tradicionais de VI são utilizadas na recuperação de informação. A seguir serão descritos alguns modelos com o objetivo de situar o contexto da VI na exibição de informações. Neste cenário, Valdivia (2007) explorou algumas técnicas e ferramentas de visualização de dados para mapeamento visual de textos. O Quadro 4 lista as principais ideias:

Quadro 4: Ferramentas de visualização

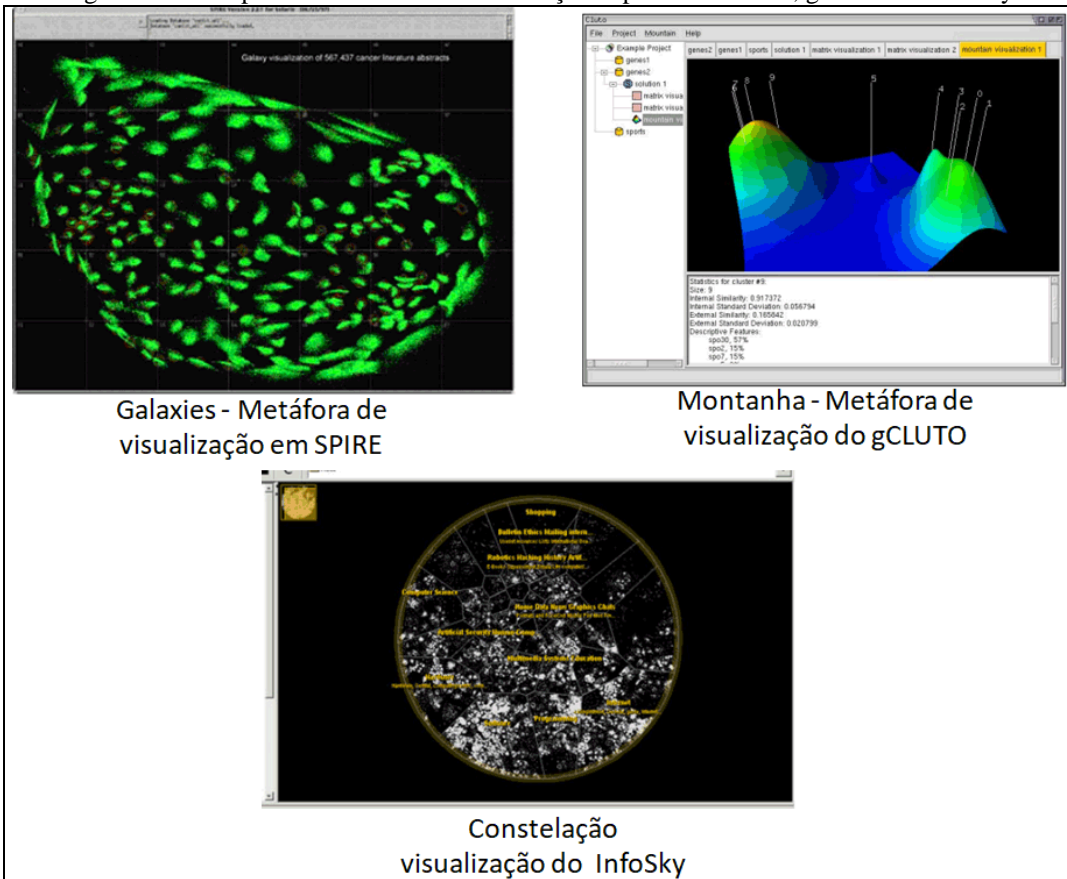
TÉCNICA/FERRAMENTA	DESCRIÇÃO
IN-SPIRE (<i>Spatial Paradigm FOR Information Retrieval</i>)	Mostra graficamente a similaridade existente entre um conjunto de textos. Agrupa textos similares próximos num espaço de visualização bidimensional em que os textos são apresentados como estrelas e os agrupamentos entre si como constelações. Essa forma de representação permite visualizar globalmente os temas e explorar a informação de interesse. Exemplo em: https://in-spire.pnnl.gov/
<i>Self-Organization Maps</i> (SOM)	O SOM é provavelmente o algoritmo mais popular usado no mapeamento de textos. Ele faz uso de algoritmos de redes neurais para gerar mapas auto-organizáveis de informação. Na superfície de visualização os textos são organizados em grupos, que são rotulados ao calcular os 3 termos com maior peso e que representam o tema principal dos textos agrupados. Exemplo em: http://www.ai-junkie.com/ann/som/som5.html
Et-Map (<i>ENTERTAINMENT MAPS</i>)	No Et-map o espaço visual é dividido em regiões, cada região possui uma cor diferente e é rotulada com um título que resume um conjunto de páginas de conteúdo similar, assim como o número de textos que pertencem a esse grupo.
Gcluto	Pacote desenvolvido para agrupar conjuntos de dados de diversos domínios do conhecimento. No Gcluto existem duas formas de visualizar dados multidimensionais: visualização da matriz de dados e visualização de montanha. A visualização da matriz apresenta uma escala de cores para representar graficamente os dados originais com poucas alterações. Na visualização por montanha, cada grupo representa um pico 3D. A localização do pico, o volume, a altura e as cores são usados para refletir informações sobre o grupo a que se referem.
Infosky	O sistema para exploração visual Infoksy emprega a metáfora de visão de uma galáxia, onde as estrelas representam textos e as constelações representam coleções de textos. A representação gráfica é realizada em um espaço de duas dimensões. As coleções de textos estão delimitadas por polígonos, de forma similar à representação de constelações em mapas estelares.

Fonte: Elaborado pelo autor baseado em Valdivia(2007)

As ideias elencadas no Quadro 4 apenas exemplificam e demonstram a possibilidade de utilização de técnicas que permitam a visualização e exploração de resultados de busca de documentos. Em resumo, utilizam-se metáforas do mundo real para representar e agrupar os documentos e tanto o IN-SPIRE como o InfoSky utilizam-se de representações de estrelas e constelações para definir conjuntos de textos. Já o gCLUTO e Vxinsight apresentam metáfora de relevos em que montanhas e picos apresentam informações sobre o grupo de textos. No SOM também são agrupados os resultados e são apresentados os 3 termos com maior peso e que representam o tema principal dos textos. Um exemplo deste algoritmo foi ilustrado na Figura 11. A Figura 12 demonstra algumas dessas metáforas.

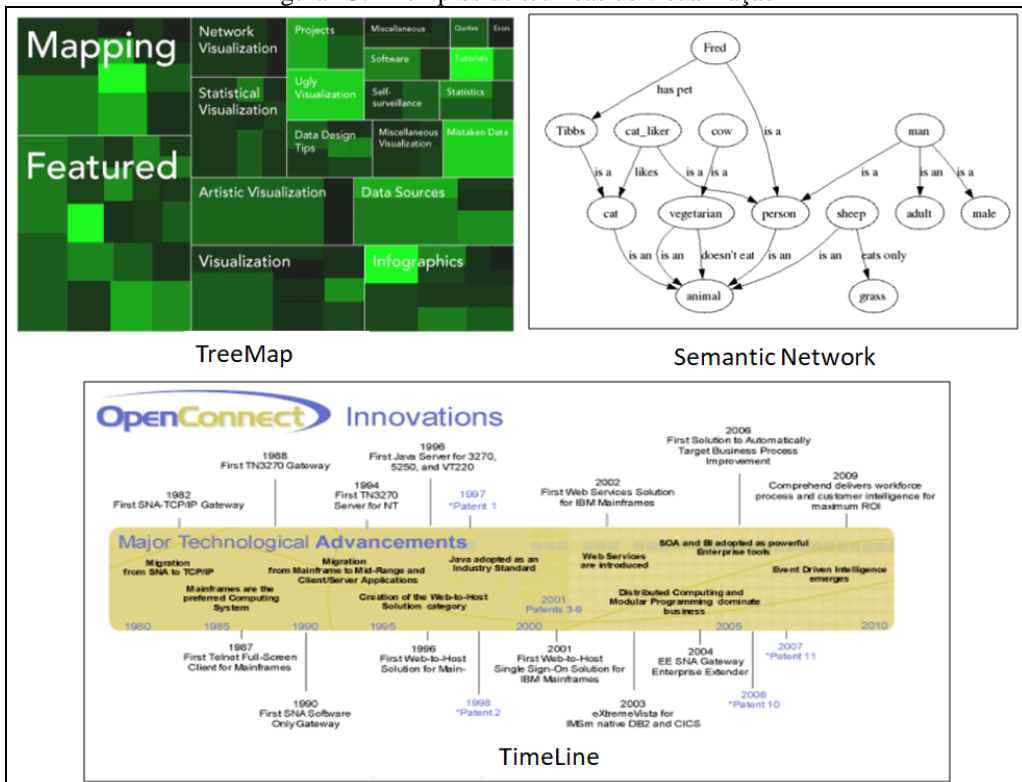
Khan e Khan (2011) discutiram as seguintes técnicas convencionais de visualização de dados: *Tree Map, Entity Relationship Diagram, Time Line, Flow Chart, Data Flow Diagram, Venn Diagram e Semantic Network*. A Figura 13 apresenta exemplos.

Figura 12: Exemplos das técnicas de visualização exploradas SPIRE, gCLUTO e Infosky



Fonte: Valdivia (2007, p.31-34)

Figura 13: Exemplos de técnicas de visualização



Fonte: Khan e Khan (2011, p.8-10)

O intuito dos pesquisadores foi apresentar as técnicas básicas de visualização para futuras pesquisas sobre VI. Resumidamente são destacadas a utilização de cada técnica para situações específicas como o *Tree Map* que é uma técnica usada para exibir dados hierárquicos na forma de retângulos aninhados ou em camadas. Já a técnica de *Time Line* é utilizada para apresentação de eventos e dados em ordem cronológica. Alguns modelos diagramáticos foram ilustrados como o *Flow Chart*, utilizado para representação simbólica de etapas envolvidas em um processo, o *Entity Relationship Diagram*, utilizado na Engenharia de Software para apresentar relações de entidades, e os diagramas de Venn que explicam a relação entre dois ou muitos conjuntos. Foi descrito o *Semantic Network* como uma rede semântica para representação gráfica da lógica da relação entre diferentes conceitos que são representados através de gráficos combinando nós ou vértices, arestas ou arcos com rótulos sobre cada aresta.

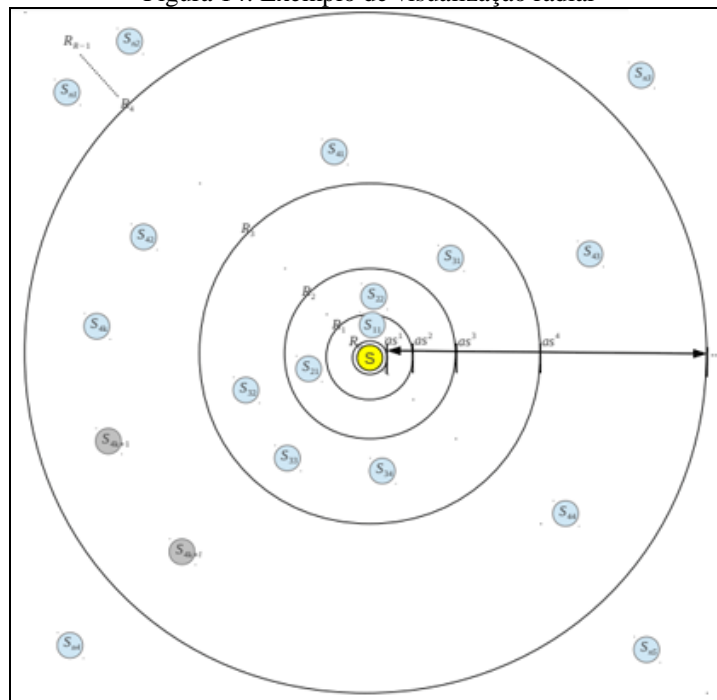
Khan e Khan (2011) concluem que os métodos de visualização são considerados muito importantes para os usuários porque fornecem modelos mentais da informação; além disso as técnicas de visualização tornam informações complexas mais compreensíveis. Nas técnicas estudadas, diferentes estratégias foram introduzidas para produzir uma visão mais ampla da informação e para a navegação através dela.

Ainda na área de visualização de informação, Marcos (2005) buscou situações que visam oferecer soluções para a representação da informação na recuperação de informação. Foram verificadas as seguintes técnicas e ferramentas: Self organizing maps (SOM), VxInsight, Cat-a-Cone, BEAD, IN-SPIRE, JAIR, NIRVE, PRISE, Perspective Wall e VIBE. Neste trabalho foram identificadas as técnicas que objetivam tornar o processo de recuperação mais compreensível e mais interativo. Dentre elas destaca-se o uso de ícones para representar conceitos, e de cores e texturas para destacar ou diferenciar elementos. A autora explica que a utilização de gráficos para apresentações hierárquicas é mais intuitiva, além de evidenciar a aplicação de efeitos de zoom para mostrar informações detalhadas. O estudo concluiu que a VI pode proporcionar modelos que minimizam o esforço mental na busca por informação, mas que em geral novas perspectivas de visualização são protótipos feitos em laboratórios de pesquisa. Por fim, constata-se que é necessário continuar desenvolvendo aplicativos que melhorem a comunicação do sistema com o usuário, em última instância, a Interação Humano Computador.

Hoffman (1997) descreveu técnicas de exploração e visualização de dados. Apresentou o Radviz como uma técnica de visualização multidimensional não linear que pode exibir dados definidos por três ou mais variáveis em uma projeção bidimensional. As

variáveis visualizadas são apresentadas como pontos de ancoragem, igualmente espaçados ao redor do perímetro de um círculo unitário. As instâncias de dados são dispostas como pontos igualmente espaçados ao redor do perímetro de um círculo. Antes da visualização, os valores das variáveis são escalonados para ficar entre 0 e 1. No centro do círculo unitário fica a variável de dados principal; os dados mais similares estão dispostos mais próximos do círculo unitário. Isto representa uma transformação não linear dos dados que preservam certas simetrias e que produzem uma exibição intuitiva. A Figura 14 exemplifica este tipo de visualização radial concêntrica.

Figura 14: Exemplo de visualização radial



Fonte: Adami (2015, p.41)

Esta seção conjecturou pesquisas que discutiram técnicas de visualização de informação. Pôde-se observar que algumas situações são utilizadas em casos de dados específicos como o caso da pesquisa de Khan e Khan (2011). Entretanto, a partir de uma visão mais geral verifica-se que todas as pesquisas apontam para a utilização de formas gráficas e metáforas do mundo real para proporcionar formas mais intuitivas de aquisição do conhecimento. Conforme discutido na seção 3.4, os estudos demonstram certa resistência à utilização de novas formas de visualização, entretanto indicam a necessidade de utilizar determinados recursos para melhorar a eficiência no processo de busca. Esses fatores são respaldados pelo pensamento peirceano que considera que a mente humana percebe de maneira mais eficiente informações diagramáticas. Esta questão será considerada no capítulo

seguinte que trata da semiótica do filósofo Charles Sanders Peirce.

3.6 SISTEMATIZAÇÃO E DISCUSSÃO

Este capítulo teve como meta evidenciar a área de IHC com o intuito de fornecer elementos a serem utilizados na proposta de interface desta pesquisa. O breve histórico se mostrou fundamental para identificar autores que contribuíram para o surgimento e a evolução da área. Os conceitos de interfaces e interação apresentam conexões importantes no apoio ao desenvolvimento de interfaces de resultados de busca nos sistemas de recuperação de informação. A conceituação de interface esclarece a importância do conjunto de equipamentos (*hardware*) e os programas (*softwares*) que proporcionam a interação do usuário com as possíveis informações disponíveis. No que diz respeito ao conceito de interação e os seus estilos, verifica-se que a combinação desses estilos pode proporcionar resultados relevantes a serem implementados nessa proposta.

A Engenharia Cognitiva e a Engenharia Semiótica são as bases que se complementam para proporcionar sustentação teórica às pesquisas voltadas à IHC. É a partir desses estudos que os usuários são considerados como ponto central no desenvolvimento de interfaces computacionais. No caso da cognitiva, o foco é disponibilizar formas mais intuitivas de executar comandos e funções. A Engenharia Semiótica considera a interação com os dispositivos como uma forma de comunicação intermediada pelo computador. Dessa forma os elementos a serem utilizados são projetados considerando essas vertentes. Busca-se proporcionar a facilidade de utilização de sistemas a partir de componentes que permitam o processo de comunicação entre projetista e usuários.

Através do levantamento de produções científicas recentes foi possível diagnosticar o panorama atual das pesquisas relacionadas e direcionamentos a serem considerados. Por fim, a IHC foi associada à área de Visualização de Informação que dispõe de perspectivas consideráveis em relação a formas e técnicas de visualização que tendem a facilitar a percepção de informações em sistemas automatizados. Com base nas pesquisas voltadas às técnicas de visualização de informação, revelaram-se perspectivas importantes a serem consideradas, principalmente no que tange à utilização de elementos visuais e diagramáticos na apresentação da informação recuperada. Concluindo, o desenvolvimento deste capítulo propiciou o estudo de conceitos e ideias a serem utilizadas na proposta de interface.

4 SEMIÓTICA PEIRCEANA PARA O DESENVOLVIMENTO DE INTERFACES

Neste capítulo está apresentada uma revisão conceitual dos elementos da semiótica peirceana que podem contribuir para o entendimento dos fatores intervenientes da Recuperação da Informação em interfaces digitais. Busca-se apresentar a ideia do raciocínio diagramático de Charles S. Peirce como apoio à representação visual. Também procura-se contextualizar a teoria da percepção de Peirce junto a outras teorias que envolvem a percepção, além de examinar as aplicações dessas teorias na área da Ciência da Informação. Conforme explica Santaella (2012b), o entendimento da teoria da percepção de Peirce está intimamente ligada aos conceitos da sua arquitetura filosófica. Presume-se que a semiótica peirceana explica o processo perceptivo com base em uma rede de semiose, contribuindo para o entendimento do processo de percepção. Além disso, outro ponto, é que os fenômenos se apresentam por meio de signos que são percebidos e interpretados e podem influenciar o processo de percepção no instante da recuperação da informação.

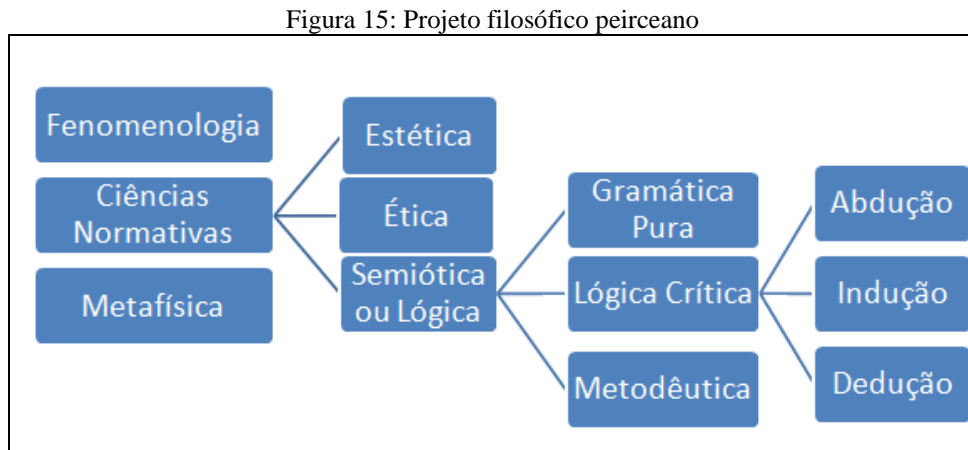
4.1 SEMIÓTICA PEIRCEANA

Algumas das bases conceituais para a elaboração deste trabalho foram a fenomenologia e a semiótica peirceana. Com isso, faz-se essencial a introdução desses temas, ainda que sucinta, para o desenvolvimento do assunto. Dentre os diversos conceitos que compõem essas áreas, objetiva-se realçar os elementos que servem como base para a teoria da percepção de Peirce.

Diferentes autores exploram a semiótica sob várias perspectivas. Em a "Semiótica do Século XX", Nöth (1996) abordou alguns desses teóricos como Saussure, Hjelmslev, Jakobson, Barthes, Eco, entre outros. Embora não tenha abordado Peirce junto aos teóricos do século XX, ele reconhece a importância dos paradigmas peirceanos que foram tratados em uma obra anterior do autor. Apesar de existirem diversas abordagens, limita-se à definição de semiótica empregada por Santaella (1983 apud NÖTH, 2003, p. 17), que se alinha com o pensamento peirceano. Neste ponto de vista, a "semiótica é a ciência dos signos e dos processos significativos (semiose) na natureza e na cultura".

Na arquitetura filosófica de Charles Sanders Peirce há um caminho importante que liga a fenomenologia ao pragmatismo percorrendo as ciências estética, ética, lógica ou semiótica e metafísica.

No entanto, é fundamental alinhar essas teorias ao escopo desta pesquisa. Reconhecendo que nem a semiótica nem as outras ciências mencionadas são um método, este alinhamento deve estabelecer-se através dos princípios do pensamento peirceano que, além de filosófico, é, sobretudo, lógico. Santaella (2012a, p.41) apresenta a seguinte estrutura do projeto filosófico peirceano (Figura 15):



Fonte: Adaptado de Santaella (2012a, p.41)

Conforme descrito por Santaella (2012a, p. 44), apesar da fenomenologia se desenvolver independente, ela serviu como alicerce para as ciências normativas. A Figura 15 demonstra o panorama geral da filosofia de Peirce. Com base nela serão destacados os componentes que irão contribuir para a compreensão da sua teoria da percepção. Inicialmente será apresentada a fenomenologia e suas categorias universais que são a base para os outros elementos. Os outros dois componentes estão presentes na semiótica: a gramática especulativa com a concepção de signos e seus desdobramentos, e a lógica crítica, através dos conceitos de abdução, indução e dedução.

4.1.1 Fenomenologia

Na visão de Santaella (2012a), a preocupação fenomenológica constitui a base de toda filosofia de Peirce. A fenomenologia teria como objetivo postular formas e propriedades universais observando os fenômenos em qualquer área da ciência. A palavra fenômeno deriva de *phaneron*, ou seja, qualquer coisa que aparece, de qualquer modo, à mente, sendo que na concepção de Peirce, os fenômenos não possuem nenhuma moldura preestabelecida, e ainda:

Fenômeno é qualquer coisa que aparece à mente, seja ela meramente sonhada, imaginada, concebida, vislumbrada, alucinada... Um devaneio, um cheiro, uma ideia geral e abstrata da ciência... Enfim, qualquer coisa (Santaella, 1995, p. 16).

Nöth (2003) explicou que na busca por definir categorias que servissem de modelo para conter a multiplicidade de fenômenos do mundo, Peirce desenvolveu uma fenomenologia recorrendo a três categorias que denominou *Firstness*, *Secondness* e *Thirdness*, que podem ser traduzidas por primeiridade, secundidade e terceiridade.

Primeiridade é a categoria das coisas, sem nenhuma relação com outros fenômenos do mundo [...]. Secundidade começa quando um fenômeno primeiro é relacionado a um segundo fenômeno qualquer [...]. Terceiridade é a categoria que relaciona um fenômeno segundo a um terceiro [...]. A base do signo é, portanto uma relação triádica desses três elementos (Nöth, 2003, p. 63-64).

Reforçando esta visão: "Primeiridade é o começo, aquilo que tem frescor, é original, espontâneo, livre. Secundidade é aquilo que é determinado, terminado, final, correlativo, objeto, necessitado, reativo. Terceiridade é o meio, devir, desenvolvimento" (Santaella, 2012b, p.79). A noção de signo corresponde à terceira das categorias fenomenológicas. Peirce (2005, p. 46) sintetiza que "um signo, ou representâmen, é aquilo que, sob certo aspecto ou modo, representa algo para alguém. [...] Ao signo assim criado denomino interpretante do primeiro signo. O signo representa alguma coisa, seu objeto".

Além das categorias fundamentais de Peirce e da noção de signo, outro elemento relevante ao estudo da percepção é a concepção de objeto (objeto dinâmico e imediato). Nöth (2003, p.67) fundamenta o conceito de objeto na obra de Peirce, salientando que ele pode ser "uma coisa material do mundo", do qual temos um "conhecimento perceptivo", mas também pode ser entidade meramente mental ou imaginária. Além disso, Peirce descreveu dois tipos de objeto: objeto dinâmico e objeto imediato. Santaella (2015, p. 34) esclarece que objeto imediato denota um objeto dinâmico e, portanto:

[...] o melhor caminho para começar a análise da relação objetual é o do objeto imediato. Afinal, parece não haver outro modo de começar, visto que o objeto dinâmico só se faz presente, mediatamente, via objeto imediato, este interno ao signo.

A divisão dos objetos do signo em dinâmico e imediato mostra que, com o objeto dinâmico, Peirce (1994, CP 5.212) identificou aquilo que está fora da cadeia sónica, aquilo que algumas vezes ele chamou de "real" ou "realidade", mas que pode ser também fictício. Esse esclarecimento será importante para a elucidação dos conceitos de percepto e percipuum que serão abordados adiante.

Nöth (2003, p.76) explica que "Peirce desenvolveu uma tipologia elaborada de signos com base em uma classificação do representâmen, objeto e interpretante, cada uma em três classes denominadas tricotomias". Sabendo-se que o termo tricotomia se refere à divisão dos elementos em três partes, cada uma das tricotomias listadas possuiu termos para designar esses elementos que são relacionados às categorias universais (primeiridade, secundidade e terceiridade).

Almeida (2009) descreve as três tricotomias do signo mais desenvolvidas. A primeira tricotomia parte da relação do signo com ele mesmo e está constituída por três tipos: o qualissigno, o sinsigno e o legissigno. A segunda tricotomia propõe a relação do signo com o objeto e está organizada em ícone, índice e símbolo. Por sua vez, a terceira tricotomia trata dos signos oriundos da conexão entre signo e interpretante e os três elementos que a compõem são: o rema, o dicissigno e o argumento.

Mesmo reconhecendo a importância das três tricotomias, especificamente neste trabalho, será evidenciada a segunda tricotomia (relação do signo com o objeto) que servirá como apoio à reflexão sobre percepção. Outro fator importante destacado por Nöth (2003, p.78) é que Peirce considera a segunda tricotomia como "a divisão mais importante dos signos". Os elementos que compõem esta divisão são o ícone, o índice e o símbolo. A seguir, remete-se à definição de Peirce:

Uma progressão regular de um, dois, três pode ser observada nas três ordens de signos, Ícone, Índice e Símbolo. O Ícone não tem conexão dinâmica alguma com o objeto que representa; simplesmente acontece que suas qualidades se assemelham às do objeto e excitam sensações análogas na mente para a qual é uma semelhança. Mas, na verdade, não mantêm conexão com elas. O Índice está fisicamente conectado com seu objeto; formam, ambos, um par orgânico, porém a mente interpretante nada tem a ver com essa conexão, exceto o fato de registrá-la, depois de ser estabelecida. O Símbolo está conectado ao seu objeto por força da ideia da mente que usa o símbolo, sem a qual essa conexão não existiria. (Peirce, 2005, p. 73)

Assim, por ícone entende-se que o mesmo representa algo através de uma relação de analogia, ou seja, apresenta características que se assemelham com o objeto. Exemplificando: em interfaces digitais de recuperação de informação, usualmente utiliza-se imagens, como por exemplo, a de um teclado de computador para representação do próprio objeto, entretanto ressalta-se que as semelhanças não são necessariamente visuais; um som ou um perfume poderiam ser ícones em potencial. Já o índice, por sua vez, é um indicador de um objeto, é um signo que através de indícios (causa) remete ao que está sendo representado, induzindo uma reação a quem o vê. Por fim, o símbolo apresenta uma relação de convenção

com o objeto que representa. Geralmente é estabelecido de forma arbitrária, como por exemplo uma palavra utilizada para representar um objeto qualquer.

Neste momento é pertinente salientar que a progressividade de atuação dos três elementos (ícone, índice e símbolo) dá suporte a perspectivas no que diz respeito à análise do julgamento perceptivo (terceiro elemento do modelo triádico), entretanto esta relação será apresentada nas próximas seções.

O último componente da arquitetura filosófica de Peirce a ser abordado é a lógica crítica, que também faz relação com o julgamento perceptivo através das inferências abduativas que são derivadas do conceito de abdução. Antes de Peirce, somente a indução e a dedução eram consideradas tipos ou métodos de raciocínio, tendo ele criado o termo abdução. Peirce classificou esses tipos dentro de suas categorias universais, ou seja, a abdução estaria para a primeiridade, a indução para a secundidade e a dedução em nível de terceiridade. Ibr (2006, p.93) utiliza o próprio Peirce para explicar "a abdução como o processo para formar hipóteses explicativas. É a única operação lógica a introduzir ideias novas; pois que a indução não faz mais que determinar um valor, e a dedução envolve apenas as consequências necessárias de uma pura hipótese" (Peirce, 1994, CP 5.171, p. 1736).

4.1.2 Peirce e o Raciocínio Diagramático

Segundo Hoffmann (2013), o raciocínio diagramático explicado por Peirce tem como finalidade simplificar os processos de pensamento tanto individuais como sociais. Este conceito pode ser abordado através de pontos de vista distintos. Optou-se, neste momento, por analisar sob a perspectiva da percepção visual. Propõe-se que esta vertente forneça suporte à ideia de que a apresentação de elementos visuais na interface são uma forma adequada de proporcionar aos usuários experiências visuais que se aproximem do modo como raciocinam e constituem seus processos cognitivos.

A ideia de combinar elementos visuais à linguagem escrita para representar documentos recuperados em uma interface computacional está apoiada no entendimento de Peirce: "raciocinar, é uma espécie de ação que se organiza através da diagramação de conceitos e inferências que ela aciona" (SANTAELLA, 2005, p.348). Consequentemente, propõe-se facilitar os processos cognitivos dos usuários ao apresentar os documentos recuperados em formato de diagrama, ou seja, mais próximo ao modo como os indivíduos fazem suas conexões mentais. Franco e Borges (2017) esclarecem que na obra de Peirce o conceito de diagrama é classificado como um tipo de ícone que representa por semelhança e

também é considerado uma máquina formal para o raciocínio. Torna-se relevante apresentar estudos que corroboram esse entendimento e apresentam reflexões acerca da percepção visual e também sobre o raciocínio diagramático.

No livro *Information Visualization Perception for Design*, Ware (2004) discorre sobre a semiótica ponderando que linguagens visuais (não escritas) são mais fáceis de aprender e utilizar. Através de exemplos de uma pintura rupestre, um diagrama esquemático e uma equação matemática o autor apresenta a necessidade de se conhecer os conceitos que as linguagens carregam para se expressar, e que em um primeiro momento possivelmente os seres humanos teriam mais experiência com os conceitos expressos nas pinturas rupestres e menos na notação matemática, obviamente se referindo a pessoas não iniciadas nessa área.

Marcondes Filho (2013) discutiu o método diagramático de Peirce estabelecendo uma relação da ideia de diagrama com a de ícone, observando uma exibição por similaridade da informação a ser representada. Na visão dos semióticos peirceanos o raciocínio diagramático sustentado na representação visual torna mais fácil ao ser humano inferir visualmente em diagramas em comparação a fórmulas lógicas ou linguagens escritas.

A conceituação de diagrama, segundo Farias (2012), é a expressão da associação entre elementos utilizando relações em um ambiente visual. O autor explica que Peirce habitualmente destacou o importante papel dos diagramas no raciocínio, denominando esse aspecto como "raciocínio diagramático". Peirce (1994, CP 4.571, p.1526) afirma que "o raciocínio diagramático é o único tipo realmente fértil de raciocínio". Nesta afirmação observa-se que Peirce reconhece a eficiência de um tipo de pensamento eminentemente visual, fundamentado na composição e utilização de diagramas.

Conforme analisado por Inácio (2006), a ideia diagramática tem influenciado várias áreas da Ciência da Computação, em especial na representação do conhecimento, design de sistemas, programação visual e desenvolvimento de interfaces. O raciocínio diagramático fundamenta esta perspectiva que considera os diagramas no auxílio ao processamento de informação através de especificações visuais de proximidade, forma, cor e padrões. O autor cita algumas vantagens do raciocínio diagramático:

- a) a facilidade da localização dos problemas devido às especificações visuais;
- b) a menor necessidade de símbolos ou etiquetas, consequência precisamente do ponto anterior;
- c) é mais fácil ao ser humano inferir visualmente em diagramas, demorando menos tempo do que em processos de fórmulas lógicas;

d) as inferências estão já presentes no diagrama: propriedades emergentes.

Em resumo, os diagramas são elementos visuais de apresentação de dados e conhecimento. Na visão de Peirce, o pensamento diagramático é uma ferramenta para produzir conhecimento (HOFFMANN, 2013). A semiótica peirceana evidencia alguns elementos relevantes e basilares no que tange ao raciocínio humano. Os conceitos da semiótica e as ideias sobre o raciocínio diagramático são essenciais para a próxima seção. A seguir será exposta a teoria da percepção de Peirce e as possíveis relações com o desenvolvimento de interfaces digitais.

4.2 TEORIA DA PERCEPÇÃO EM PEIRCE E AS CONTRIBUIÇÕES PARA O DESENVOLVIMENTO DE INTERFACES

O objetivo inicial desta seção é apresentar os tópicos centrais da teoria da percepção em Peirce. Em um segundo momento, será relacionada a abordagem da percepção em Peirce com o desenvolvimento de interfaces Web. Contudo, inicia-se por tratar o assunto de maneira geral. As pesquisas sobre percepção têm colaborado para ampliar a compreensão de cognição, no sentido de analisar como se realiza o fenômeno "conhecer". Santaella (2012b, p. 1) inicia sua obra sobre percepção enfatizando o crescente interesse sobre o tema, sobretudo a partir do século XIX, com o impulso das ciências cognitivas. Essas pesquisas estão inseridas em diferentes abordagens que permeiam as diversas áreas de conhecimento.

Uma das formas de se introduzir um tema é a partir de uma definição, porém a palavra percepção, conforme discorreremos durante esta seção, não se refere a um evento isolado. Existem diversas teorias que abordam o tema como a Empirista, a Intelectualista, a Psicologia da forma (ou teoria da Gestalt) e a Fenomenologia, em que a percepção é descrita através de um processo integrado e pode ser considerada fundamental para compreender o conhecimento humano. Desta forma, entende-se que a percepção é "a ponte entre o mundo da linguagem - a consciência, o cérebro, a mente - e o mundo lá fora" (Santaella, 2012b, p. 75).

Para Barros e Café (2012) os processos de significação são importantes aspectos da Ciência da Informação (CI) no que diz respeito à organização e representação do conhecimento, uma vez que a CI desenvolve métodos para interpretar diferentes constituições textuais. Para as autoras, esse processo de significação também é vislumbrado pela Semiótica, ciência dos signos produzidos nos âmbitos da natureza e da cultura (Santaella, 1983 apud

Nöth, 2003). O processo de percepção pode ser explicado pela arquitetura filosófica de Charles Sanders Peirce (1839-1914). Neste sentido, verifica-se a possibilidade de buscar aspectos da semiótica peirceana que possam contribuir para compreensão do processo de percepção humana.

Antes, julgou-se necessário analisar algumas produções que versam sobre as teorias da percepção e suas aplicações, com o objetivo de contextualizar o tema. Durante a busca por trabalhos referentes ao tema foram analisados os resumos de diversos trabalhos, verificou-se que a maioria das pesquisas sobre percepção apresentou alguma ligação com o sistema visual humano, tais como: percepção visual, visualização de informação e ensino (Correia, 2001; Loomis, 2003; Santos, 2003; Alexandre e Tavares, 2007; Sousa, 2009; Andrade et al., 2012; Jorge; Rezende e Wartha, 2013; Matsunaga, 2015).

Outras pesquisas apresentam análises individuais das teorias da percepção de Peirce, Merleau-Ponty e Gödel (Frankenthal, 2004; Nobrega, 2008; Schultz, 2012). Rego, Pessoa e Gala (2015) discutiram as diferenças da teoria da percepção de Bertand Russell, Charles Sanders Peirce e Alberto Caeiro. Moraes, Pereira e Pantaleão (2012) analisaram as correlações da teoria peirceana para a abordagem ecológica através da aproximação entre as perspectivas ecológica e pragmatista para verificar em que medida o pragmatismo peirceano contribuiria para o estudo da percepção. Os autores relataram que a partir dos pressupostos do realismo direto, percepção direta (em certo grau) e abordagem sistêmica, é possível a aproximação entre as abordagens classificando como um novo campo conceitual para a compreensão da percepção, segundo a perspectiva direta. A única ressalva é quanto ao julgamento perceptivo (elemento essencial à percepção, segundo Peirce) , pois "extrapola o escopo da abordagem ecológica, uma vez que engloba regras e hábitos, que podem ser simultâneos à percepção, que não são explicitados pela abordagem ecológica" (Moraes, Pereira e Pantaleão, 2012, p. 203).

Algumas produções científicas exemplificaram a aplicação da teoria da percepção de Peirce em espaços geográficos (Rocha, 2002; Mucelin e Bellini, 2008). Enfim, estes exemplos proporcionam uma visão mais ampla sobre o conceito de percepção e seus potenciais de aplicação. Além disso, ratificaram a importância dos comentadores e teóricos da semiótica e da percepção.

Nos diversos temas apresentados, invariavelmente surgiram questões como: "O que é percepção?", e "Qual o significado de percepção?" Uma das primeiras ideias seria a busca do termo em um dicionário. Entretanto, quando se realiza um estudo um pouco mais minucioso, é possível constatar que este tipo de significação é bem restrito, já que sua

definição possui diferentes conceitos, a depender da abordagem teórica. As abordagens sobre percepção envolvem uma variedade de compreensões que nem sempre dirigem-se para um consenso. A seguir, será apresentado um panorama dessas abordagens.

Para prosseguir com a inquirição sobre o conceito de percepção, é importante confrontá-lo com o conceito de sensação. As teorias que discutem essa relação são apresentadas no Quadro 5.

Quadro 5: Teorias sobre sensação e percepção

Teorias	Concepções
Empirista	<p>A sensação e a percepção dependem das coisas exteriores;</p> <p>A sensação seria pontual, isto é, um ponto do objeto externo toca um de meus órgãos dos sentidos e faz um percurso no interior do meu corpo, indo ao cérebro e voltando às extremidades sensoriais;</p> <p>Cada sensação é independente das outras e cabe à percepção unificá-las e organizá-las numa síntese;</p> <p>A causa do conhecimento sensível é a coisa externa, de modo que a sensação e a percepção são efeitos passivos de uma atividade dos corpos exteriores sobre o nosso corpo;</p> <p>O conhecimento é obtido por soma e associação das sensações na percepção e tal soma e associação dependem da frequência, da repetição e da sucessão dos estímulos externos e de nossos hábitos.</p>
Intelectualista	<p>A sensação e a percepção dependem do sujeito do conhecimento e a coisa exterior é apenas a ocasião para que tenhamos a sensação ou a percepção;</p> <p>Sentir e perceber são fenômenos que dependem da capacidade do sujeito para decompor um objeto em suas qualidades simples (a sensação) e de recompor o objeto como um todo, dando-lhe organização e interpretação (a percepção);</p> <p>A passagem da sensação para a percepção é, neste caso, um ato realizado pelo intelecto do sujeito do conhecimento, que confere organização e sentido às sensações;</p> <p>Não haveria algo propriamente chamado percepção, mas sensações dispersas ou elementares; sua organização ou síntese seria feita pela inteligência e receberia o nome de percepção.</p>
Psicologia da forma e fenomenologia	<p>Contra o empirismo, que a sensação não é reflexo pontual ou uma resposta físico-fisiológica a um estímulo externo também pontual;</p> <p>Contra o intelectualismo, que a percepção não é uma atividade sintética feita pelo pensamento sobre as sensações;</p> <p>Contra o empirismo e o intelectualismo, que não há diferença entre sensação e percepção.</p>

Fonte: Elaborado pelo autor baseado em Chauí (2000, p. 151-153)

A partir dessas informações, Chauí (2000) conclui que para os empiristas e intelectualistas a percepção era considerada a atividade que unia as partes numa síntese que seria o objeto percebido. Na psicologia da forma e na fenomenologia a autora salienta que

[...] não há diferença entre sensação e percepção porque nunca temos sensações parciais, pontuais ou elementares, isto é, sensações separadas de cada qualidade, que depois o espírito juntaria e organizaria como percepção de um único objeto. Sentimos e percebemos formas, isto é, totalidades estruturadas dotadas de sentido ou de significação. (Chauí, 2000, p. 153).

Pode-se constatar que existem várias abordagens em relação à sensação e

percepção, mas independentemente da perspectiva, são temas relacionados e, por muitas vezes, estudados em conjunto. A seguir, apresentam-se as abordagens de Merleau-Ponty e Gibson.

Merleau-Ponty, em sua teoria da Fenomenologia da Percepção, apresenta uma análise da percepção por meio da revisão do conceito de sensação. Santaella (2012b) observa que para Merleau-Ponty deve ser rejeitada qualquer tentativa de decomposição da percepção em sensações. Seguindo a mesma linha, na teoria de Gestalt a mente humana configura as informações através dos canais sensoriais, percepção e/ou da memória (pensamento, cognição e resolução de problemas). Na experiência humana do meio ambiente, esta configuração tem um caráter primário sobre os elementos que o compõem, e assim a soma desses elementos por si só não poderia levar-nos à compreensão do todo.

Neste panorama, das teorias que versam sobre percepção, é importante citar "A Ecologia da Percepção", cujo fundador foi James J. Gibson. Seria inviável sintetizá-la neste estudo, entretanto podemos elencar algumas contribuições importantes. Para Santaella (2012b), destacam-se algumas características: designação do conceito de *Affordance*: tudo aquilo que o ambiente oferece ao observador. Esta teoria sofreu influência da Gestalt, mas também a refutou; as teses gibsonianas mais refutadas e submetidas a discussões são a percepção direta (sem a mediação de representações mentais internas) e imediata e a da negação de qualquer papel desempenhado pela mente na percepção.

Completando este cenário das teorias, posiciona-se aqui a Teoria da Percepção de Peirce que é um dos fundamentos deste trabalho.

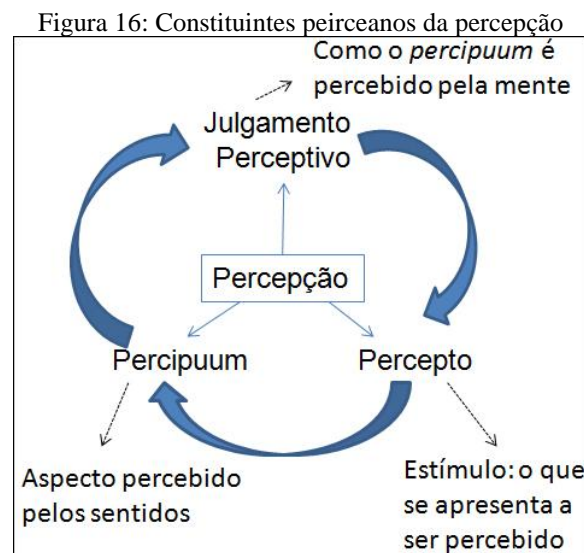
4.2.1 O processo de percepção em Peirce

Pretende-se abordar, neste item, as conexões entre a semiótica peirceana e o contexto em que essa teoria se desenvolveu, com o objetivo de fornecer uma visão clara sobre o tema. Algumas dessas conexões já foram descritas ao longo do trabalho, mas reaparecem aqui de forma mais ordenada. De acordo com Santaella (2012b), foram poucos estudiosos da obra peirceana que se empenharam em investigar seus escritos sobre percepção. Ela realizou uma análise dos comentadores de Peirce que se dedicaram ao tema, e também aos ensaios peirceanos publicados. Como consequência desta exploração, destacaram-se diversos elementos importantes que serviram de suporte ao desenvolvimento deste assunto.

No livro "A percepção: uma teoria semiótica", Santaella (1998, p. 16) afirma que "em termos lógicos, ontológicos e epistemológicos", a teoria peirceana sobre

percepção é altamente completa, Jorge, Rezende e Wartha (2013, p. 156) complementam afirmando que para Peirce "as representações cognitivas são signos, representações mentais são modelos de processos cognitivos e operações mentais ocorrem na forma de processos sígnicos". Nessa perspectiva, investiga-se a relevância do processo perceptivo na construção do conhecimento, sendo a percepção o elo entre a mente e o mundo exterior. Com o objetivo de esclarecer a composição da teoria da percepção, apresenta-se a seguir o modelo triádico formulado por Peirce e que está intimamente conectado aos conceitos já apresentados.

Ao iniciar a explanação sobre o modelo triádico da percepção peirceana é importante salientar que os elementos a serem descritos a seguir são associados aos conceitos iniciais do seu projeto filosófico. Na explicação de Santaella (2012b, p. 75), a percepção desempenha o papel de ponte entre o mundo da linguagem - a consciência, o cérebro, a mente – e o mundo lá fora. A percepção é, para Peirce, o objeto de estudo da semiótica que ocorre segundo um modelo triádico e está constituído de Percepto, *Percipuum* e Julgamento Perceptivo. A autora salienta que Peirce produziu a única teoria triádica da percepção, e explica que a percepção está sob a dominância da secundidade. A Figura 16 procura mostrar a relação entre os elementos peirceanos da percepção.



Fonte: Elaborado pelo autor baseado em Mucelin e Bellini (2008, p.124)

A partir da proposta triádica ilustrada pela Figura 16, Mucelin e Bellini (2008, p.125) consideram os constituintes da percepção como "entes interdependentes e indecomponíveis que permitem que se analise e caracterize isoladamente cada um deles". Explica-se esses elementos da seguinte forma: percepto ou objeto, verdadeira coisa em si mesma, que independe daquilo que dele se possa pensar, devido à sua existência e insistência

sobre os sentidos sem nada manifestar; o percipuum ou o modo como o percepto, captado pelos órgãos sensoriais, é imediatamente interpretado no julgamento de percepção; e o julgamento perceptivo, que corresponde a uma espécie de proposição a nos informar sobre aquilo que está sendo percebido. Ainda sobre essa relação triádica, deve-se destacar uma passagem do próprio Peirce:

Não sabemos nada sobre o percepto a não ser pelo testemunho do juízo perceptual, exceto que sentimos o seu golpe, a reação dele contra nós, e vemos os conteúdos dele arranjados num objeto, na sua totalidade – excetuando-se também, é claro, o que os psicólogos são capazes de extrair inferencialmente. Mas, no momento em que fixamos nossa mente sobre ele e pensamos a menor coisa sobre o percepto, é o juízo perceptual que nos diz o que nós assim ‘percebemos’. Por esta e outras razões, proponho considerar o percepto tal como ele é imediatamente interpretado no juízo perceptual, sob o nome de percipuum. (Peirce, 1994, p. 39, CP 7.643)

Peirce nos apresenta nesse trecho a indissociabilidade dos elementos no processo perceptual. Nesta concepção, considera-se o percipuum a imediata interpretação do percepto através do juízo perceptual, ou seja, o percipuum apresenta-se como um elemento derivado do próprio percepto que já é resultado de uma elaboração mental que se utiliza da organização que regula os nossos julgamentos. Santaella (2012b, p. 95) também discorre sobre a tríade perceptiva, da seguinte maneira:

A percepção é determinada pelo percepto, mas este só pode ser conhecido através da mediação do signo, que é o julgamento da percepção. Para que esse conhecimento se dê, o percepto deve, de algum modo, estar representado no signo. Aquilo que representa o percepto, dentro do julgamento perceptivo, é o percipuum, meio mental de ligação entre o que está fora e o juízo perceptivo, que já é fruto de uma elaboração mental (SANTAELLA, 2012b, p. 95).

Hausman (2006) comenta que Peirce discute a função do percepto no início do processo de interpretação. Nesta visão, são sugeridos dois novos elementos o "antecepto" e o "ponecepto". Um percepto isolado é apenas uma possibilidade perceptiva. O percepto sempre ocorre em sequências seriais que envolvem memórias recentes e antecipações futuras, chamadas respectivamente de poneceptos e anteceptos. Romanini (2006, p.68) complementa:

Perceptos similares chegam e coalescem na mente para formar o universo perceptivo, uma espécie de imagem mental, uma “ideia” que é o ground do julgamento perceptivo. Essa ideia é uma metáfora. O percipuum, portanto, é o percepto representado no julgamento perceptivo. É uma ficção, uma hipótese, assim como também o são o ponecipuum e o antecipuum.

De acordo com Santaella (2012b), ao analisar a teoria peirceana da percepção, no âmbito da semiótica, o percepto funciona como objeto dinâmico enquanto o

percipuum seria o objeto imediato. Nesta mesma análise é possível incluir o terceiro elemento da tríade, o julgamento de percepção. Santaella (2012b) referencia a obra de R. Bernstein (1964), em que o julgamento de percepção é correlacionado com as inferências abduativas, sendo o elemento que desempenha o papel cognitivo na percepção. Nesse aspecto, a autora exemplifica afirmando que "são os juízos perceptivos que nos dizem, por exemplo, que o cheiro que estamos sentindo é de brócolis cozido, que aquilo que estamos vendo é uma lua cheia solitariamente iluminando o céu" (Santaella, 2012b, p. 95).

A tríade da percepção permite vislumbrar análises de diferentes situações do cotidiano. Para reforçar o entendimento, pode-se exemplificar através da seguinte situação: um indivíduo observa uma obra de arte. Nesta breve circunstância pode-se ilustrar o funcionamento da percepção, identificando os três elementos já apresentados. Primeiramente, a obra de arte em si seria o percepto, já o percipuum seria o conjunto de aspectos emanados da obra, e por fim, o aspecto percebido pelo indivíduo que se denomina julgamento perceptivo que é produzido pela mente através de um processo inferencial da percepção.

Desta forma, relaciona-se os elementos da tríade de percepção formulada por Peirce a alguns conceitos fundamentais de sua arquitetura filosófica. Em síntese, relaciona-se o percepto ao objeto dinâmico, o percipuum ao objeto imediato e o julgamento perceptivo às inferências abduativas.

A seguir, apresenta-se a relação do julgamento perceptivo com os elementos da segunda tricotomia dos signos, o ícone, o índice e o símbolo. Ao agir como signo, o julgamento perceptivo apresenta três perspectivas equivalentes aos tipos sígnicos de Peirce: a icônica, a indicial e a simbólica. Santaella (2012b, p. 128) comenta que "nessa medida, o ingrediente icônico é justamente aquilo que dá suporte ao processo perceptivo, funcionando como substrato da ilusão, subjacente a toda percepção, de que o objeto, tal como percebido, é o próprio objeto". Na visão de Granja (2005), a mediação icônica nos faz reconhecer tanto a forma geral dos objetos (a forma circular, a forma de uma cadeira), como a forma particular do objeto imediato que se apresenta à percepção. A junção entre o geral da semelhança icônica ao particular do objeto percebido corresponde justamente à dimensão indicial do julgamento perceptivo. Na sua dimensão simbólica ou de generalidade, o juízo de percepção é falível, e pode ser reavaliado a partir de experiências futuras.

Em geral, a representação de um objeto é complexa e consiste em várias funções icônicas, indexais e simbólicas que atuam na formação do julgamento perceptivo. Os ingredientes icônicos contribuem ao apresentar características gerais que permitem a identificação de traços essenciais à formação mental do percipuum. Em outro aspecto, uma

conexão indicial também é importante para fixar a referência através dos indícios que se aproximam numa relação efetiva de seu objeto; afinal o índice é um indicador do seu objeto. A dimensão simbólica opera no terceiro nível do interpretante no qual, através das características e indícios relacionados, é formulado o pensamento na mente e este pensamento é um símbolo. Dessa forma, o julgamento perceptivo produz o que é percebido mentalmente.

Conforme destacado por Rego, Pessoa e Gala (2015, p. 105), Peirce não desenvolveu um texto que sistematizasse sua teoria da percepção, entretanto alguns intérpretes de sua obra organizaram essas ideias apresentando uma teoria muito rica. Desta forma, os elementos desta teoria estão calcados em sua arquitetura filosófica na qual os conceitos do processo perceptivo estão intimamente relacionados a todo trabalho desenvolvido no decorrer de suas pesquisas.

4.2.2 A percepção e o desenvolvimento de interfaces de Recuperação de Informação

O objetivo deste tópico é encontrar elementos da teoria da percepção de Peirce que possam beneficiar o desenvolvimento de interfaces *web*, tendo em vista o processo de recuperação da informação. A percepção humana é um fator essencial a ser considerado no projeto de interfaces digitais. Andersen (2001) salienta a importância de estudos focados na interação entre as máquinas e os seres humanos. O autor destaca o núcleo da semiótica na realização de estudos que discutam as representações e também a forma como os usuários interpretam essas representações. A seguir, são elencadas algumas análises que consideram a percepção e a semiótica em diversas vertentes, e que colaboram para o desenvolvimento de interfaces digitais.

A relação entre a percepção humana e a organização do conhecimento foi destacada por Barat (2005). Nessa pesquisa apresentam-se alguns aspectos referentes à teoria da percepção e à teoria linguística. O autor discute as vantagens das teorias e na teoria perceptual elenca três itens: a representação homogênea da primeira percepção, a visualização clara e lógica e os padrões de alto nível de processamento de dados e recuperação de informação. Na teoria linguística é diagnosticado que ela é comum, conhecida e os símbolos são compreensíveis e se comunicam diretamente e instantaneamente. A partir desses conceitos são listados também alguns pontos sobre percepção visual e recuperação de informação. Em suas conclusões, o autor reafirma a importância da compreensão da percepção no processo de gestão da informação. A apresentação do estudo fortalece a ideia de analisar a percepção como fator essencial no desenvolvimento de elementos que visam à

melhor compreensão humana.

No que diz respeito aos estudos sobre interfaces digitais relacionadas à semiótica, a área de Engenharia Semiótica (ES) é apresentada por Souza (2005) como a teoria semiótica da área de IHC (Interação Humano Computador). A autora discute em seu artigo a ES como suporte teórico no desenvolvimento de interfaces com artefatos interativos que realmente possam trazer novas perspectivas e possibilidades de design em IHC. A ES trata a IHC como um tipo específico de comunicação mediada por computador, considerando a troca de mensagens dos usuários com os designers de sistemas computacionais. Essas mensagens são o *representamen* (ou seja, o sinal perceptível que representa a mensagem) em que o próprio sistema informa aos usuários como se comunicar com ele. Em favor da engenharia semiótica defende-se que ela pode ajudar os designers a compartilharem responsabilidades com os usuários no processo comunicativo.

Também seguindo a vertente da Engenharia Semiótica, Vandi (2007) avaliou a forma como a semiótica possibilita analisar a ação prática dos usuários. Reavaliando o uso de metáforas em interfaces gráficas, o autor considera metáforas os itens destinados a dar ao usuário um contexto familiar para agir, ou seja, um escritório com lixeiras, pastas e arquivos. No artigo foram exemplificadas algumas interfaces computacionais que contêm tais itens, sendo diagnosticado que a tarefa de um bom projetista é usar esses elementos para estimular a aprendizagem e a fácil criação de hábitos de ação. Os hábitos de ação iniciam-se através da percepção. Para Peirce (1994, CP 5.212), "os elementos de todo conceito entram no pensamento lógico pelos portões da percepção", ou seja, toda informação deve entrar pelas portas da percepção, apresentando-se em forma de uma ideia ou hábito mental compartilhado numa comunidade.

Islam (2013) realizou uma revisão de literatura que abordou a percepção semiótica em interfaces de usuários no que diz respeito ao projeto e avaliação de usabilidade. Foram selecionados 65 artigos que identificaram alguns desafios da área. Segundo o autor, há a necessidade de realização de mais pesquisas ligadas ao tema com o intuito de melhorar a qualidade das publicações. O autor revela que apenas 18% dos artigos apresentaram o design de interface ou princípios de avaliação de usabilidade; cerca de 51% estavam relacionados a percepção semiótica, apresentando teorias, métodos e estruturas. Após uma verificação criteriosa nos estudos, o autor conclui que a revisão mostrou a importância da semiótica nas pesquisas interface do usuário, além disso sugere direcionar novos estudos a algumas lacunas de pesquisas identificadas, dentre elas a percepção semiótica na avaliação da usabilidade.

Embora não tenha tratado especificamente sobre interfaces digitais,

Gambarato (2013) discutiu a semiótica de Peirce no processo de design, em especial o conceito de abdução. A autora apresentou o potencial comunicativo dos signos com uma breve análise da teoria da percepção, explicando a importância do conceito de abdução criado por Peirce. O artigo destaca a capacidade de representação por meio dos signos e conclui que o processo criativo no design tem o objetivo de associar signos e gerar novos interpretantes. Esse estudo apresentou-se importante no que diz respeito à visão da teoria da percepção em processos comunicativos, visto que as interfaces também podem ser consideradas parte do processo comunicacional.

Magnani e Bardone (2006) discorreram sobre a relação do conceito de abdução e o projeto para interfaces web. Os autores apresentaram a importância de três dimensões para as inferências abduativas: a visual, a espacial e a emocional. A dimensão visual é exemplificada através do uso de cores, tamanhos e formatos. Já a dimensão espacial diz respeito à disposição dos elementos na interface, proporcionando a realização de tarefas específicas. No caso da dimensão emocional, consideram-se aspectos de alertas aos usuários a partir de caixas de diálogo, sons e cores, com significados de advertência e sucesso, por exemplo. Os autores consideram a interface como fundamental na compreensão da interação humano-computador e que esta interação pode ser melhor compreendida através do modelo inferencial. Por conta da possibilidade da interface fornecer pistas a partir das quais os usuários possam ter uma melhor percepção do ambiente, os autores consideram o processo como genuinamente abduutivo.

No que diz respeito à navegação em interfaces digitais web, Santaella (2004) identifica perfis relacionados às três categorias gerais de Peirce. Os tipos de raciocínio, abduutivo, indutivo e dedutivo, determinam diferentes percursos de leitura para três tipos de leitores (ou internautas): o errante, o detetive e o previdente. Neste cenário, o internauta errante é considerado o leitor novato, que por meio do raciocínio abduutivo navega aleatoriamente através das possibilidades a ele apresentadas. A indução é própria do internauta detetive, que está em processo de aprendizado, já que navega seguindo indícios (pistas) em busca da informação que necessita. Por sua vez, o internauta previdente é capaz de prever o resultado de suas ações na pesquisa, pois já passou pelo processo de aprendizagem e aplica o raciocínio dedutivo para avançar nos ambientes informacionais. Embora esses estilos sejam diferentes, existem pontos em comum, pois independente do itinerário os internautas recebem estímulos que despertam a percepção e a cognição. Dessa forma, é identificado o papel fundamental da percepção no processo de navegação, onde os fatores sensoriais provocam imediatas reações perceptivas em sincronia com operações mentais.

Os estudos elencados no item 4.2.2 apresentaram diversos aspectos da perspectiva semiótica, englobando a teoria da percepção de Peirce, sugerindo que sua contribuição está voltada ao desenvolvimento de interfaces digitais. Contudo, verifica-se a necessidade de aprofundamento nos estudos experimentais relativos à percepção, sendo a semiótica de Peirce certamente um caminho teórico a ser percorrido.

4.3 SISTEMATIZAÇÃO E DISCUSSÃO

Estudar a percepção e decifrar como ela age na formação do pensamento, é um ganho importante para o entendimento da cognição em qualquer área disciplinar. É relacionar os diversos fatores da arquitetura filosófica peirceana que se destacam no processo perceptivo. A proposta triádica, composta pelo percepto, o percipuum e o julgamento perceptivo, conectou-se às categorias universais, aos objetos do signo e ao conceito de abdução.

A partir da apresentação das teorias, observou-se as diversas abordagens sobre o tema, além de elencar algumas aplicações na área da Ciência da Informação que utilizam essas teorias como base para suas pesquisas. Com a revisão conceitual sobre a teoria da percepção de Peirce, verifica-se que seus conceitos fornecem subsídios para análise de diversos domínios do conhecimento e proporcionam um melhor entendimento do processo de percepção, substituindo uma relação diádica cartesiana de sujeito-objeto, para uma relação triádica do processo de percepção.

Nas diferentes áreas observadas, identifica-se os componentes da tríade perceptual, isto é, entende-se que existem elementos a serem percebidos (percepto), como esses elementos se apresentam ao julgamento de percepção (percipuum), e como tudo isto é imediatamente interpretado, ou seja, o julgamento de percepção. Além dessa apresentação dos aspectos dos signos, outra vertente destacada diz respeito ao perfil dos internautas que utilizam interfaces Web, uma vez que os aspectos perceptivos e cognitivos são influenciados por diferentes itinerários de navegação que podem ser distinguidos pelos tipos de raciocínio abdução, indutivo e dedutivo. Conforme destacado por Jorge, Rezende e Wartha (2013), a Teoria da Percepção sustenta eficientemente o desenvolvimento da habilidade viso espacial, ou a questão da carga cognitiva sobre conceitos que demandam altas habilidades viso espaciais por meio do uso de ferramentas de visualização.

O desenvolvimento de interfaces digitais deve levar em conta as fases do processo de percepção elencados por Peirce, além, é claro, dos tipos de mediação (icônica,

indicial e simbólica) que se apresentam na Interação Homem-Computador. Sem estes conhecimentos, as interfaces são apenas um mecanismo artificial e pouco eficiente de possibilitar a interação humana o que, ao fim e ao cabo, interfere negativamente na aquisição de conhecimento e na fluência no uso do sistema.

5 PROPOSTA DE INTERFACE DE RESULTADO DE BUSCA

Este capítulo está estruturado para apresentar a proposta de interface de resultados de busca fundamentada nos conceitos de Recuperação de Informação, Interação Humano Computador e Semiótica. Inicialmente o protótipo será apresentado utilizando a linguagem UML (*Unified Modeling Language*). Foram produzidos diagramas que possibilitam uma visão geral do *software* e, em seguida, descritas as tecnologias utilizadas com o intuito de fornecer princípios e possibilidades de futuras pesquisas subsequentes ao projeto.

A proposta também será exposta através de imagens do protótipo realizando a relação dos elementos dispostos com os conceitos básicos discutidos na fundamentação teórica. Serão abordados conceitos de similaridade através do modelo vetorial, interação advinda da IHC e elementos da semiótica relacionados aos conceitos de signos, inferências abduativas e navegação. Desta forma, espera-se alinhar as bases teóricas com a aplicação prática dos conceitos em um produto de *software*.

5.1 MODELAGEM DO PROTÓTIPO UTILIZANDO A UML

Na medida em que se objetiva desenvolver qualquer artefato de *software*, é necessário seguir etapas através de algum modelo ou método. Mesmo não tendo o objetivo de apresentar todo o ciclo de desenvolvimento de um *software*, é necessário expor, ainda que de forma breve, elementos que norteiam a construção da proposta de interface desta pesquisa. Neste caso será utilizada a Linguagem de Modelagem Unificada (UML) que é um modelo atual, difundido e padronizado para realizar essa apresentação.

Na produção de aplicações computacionais, o ciclo de vida é considerado por Sommerville (2008) como um conjunto de tarefas e resultados que formam um produto de *software*. Conseqüentemente, esse processo ocorre através da organização de um conjunto de atividades que formam o produto final. A parte final desta pesquisa está destinada ao desenvolvimento do protótipo da interface de resultado de busca, mas para se tornar funcional é necessário que um conjunto de ferramentas e especificações seja realizado, formando um sistema de recuperação de informação. Ainda que não seja abordado todo o ciclo de vida, a apresentação de alguns diagramas facilita a compreensão geral da estrutura.

A definição de Pressman (2006) indica que o conceito de *software* leva em consideração três ideias básicas: (1) instruções que quando executadas proveem função e

performance, (2) estrutura de dados que permitem aos programas manipular adequadamente as informações e (3) informação descritiva em meio virtual ou físico que explanam a operação e o uso dos programas. Assim, verifica-se que um *software* não se restringe somente ao programa de computador ou seus dados, mas abrange também a sua documentação. Como a ideia é disponibilizar livremente esta pesquisa e suas fontes, os diagramas possibilitam a outros pesquisadores dar continuidade ou agregar elementos à proposta.

A UML é utilizada por representar os padrões e os processos de construção de *software* devido ao dinamismo e o poder de representação de seus diagramas. Pressman (2011) explica que a UML foi desenvolvida por Grady Booch, Jim Rumbaugh e Ivar Jacobson, na década de 1990, e combinou diferentes notações concorrentes que norteavam o desenvolvimento de *softwares*. A versão atual da UML é a 2.5 e possui quatorze diagramas, entretanto, nesta pesquisa, serão exibidos dois diagramas: diagrama de casos de uso e diagrama de classes. Estes dois diagramas são suficientes para a representação do modelo conceitual e da arquitetura da proposta da interface.

Os diagramas da UML possibilitam apresentar diversas características e aspectos importantes do *software*. Rodrigues (2009) explica que cada diagrama apresenta o *software* a partir de um determinado ponto de vista; alguns dão uma visão externa mais geral enquanto outros apresentam características mais específicas. Os diagramas podem expressar fatores estáticos, dinâmicos ou comportamentais do *software*, sendo que alguns deles complementam outros, possibilitando a visualização do sistema como um todo. Serão utilizados, a seguir, os diagramas de casos de uso e de classes, sendo o primeiro responsável por apresentar o comportamento do sistema para o usuário externo e, o segundo, a estrutura interna dos objetos do sistema.

5.1.1 Diagrama de casos de uso

A principal função do diagrama de casos de uso é apresentar as principais funcionalidades do *software*. Para Pressman (2011), ele é utilizado para descrever um cenário que mostra essas funcionalidades a partir do ponto de vista das pessoas envolvidas na elaboração do *software*. Rodrigues (2009) destaca que este é o diagrama menos formal da UML, devido à utilização de uma linguagem simples que facilita a compreensão do comportamento do sistema. No processo de desenvolvimento é um dos primeiros diagramas a serem confeccionados auxiliando na verificação dos requisitos do sistema.

O diagrama de casos de uso possui três elementos básicos: os atores, os

casos de uso e as associações. Pereira (2011) descreve que o termo ator refere-se ao papel de algum elemento que interage com o sistema, podendo ser usuários, dispositivos eletrônicos ou também outros sistemas computacionais. Os atores podem interagir com um ou vários casos de uso, e o nome dado a este ator deve corresponder ao papel por ele desempenhado dentro do sistema. Os casos de usos correspondem às ações que geram uma atividade no sistema e são representados por elipses, geralmente sua função é especificada através do uso de verbos no infinitivo em conjunto com um substantivo. Por fim, as associações determinam as interações e as funções que cada ator pode realizar e são representadas por segmentos de retas, curvas ou poligonais e a utilização de setas. Podem ocorrer associações entre casos de usos. Nesta situação, representa-se uma ação do sistema acessando outra funcionalidade. A Figura 17 apresenta a notação gráfica dos três elementos.

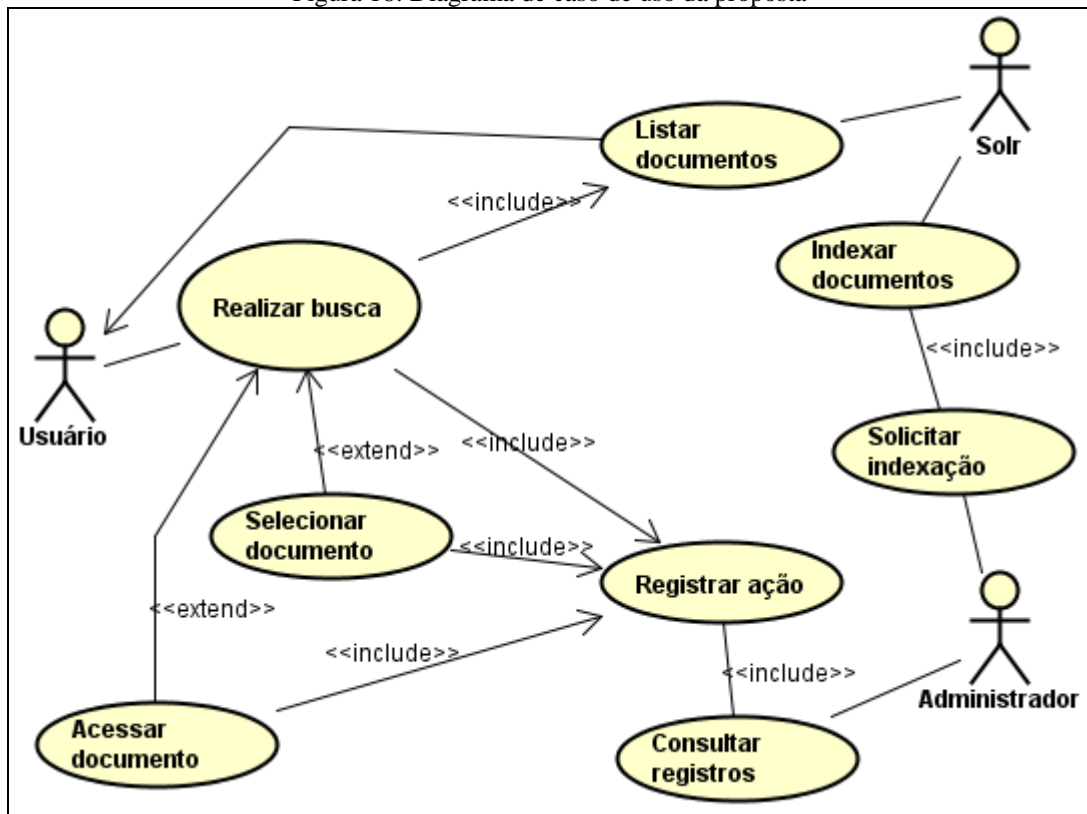
Figura 17: Elementos do diagrama de Casos de Uso



Fonte: Elaborado pelo autor

A ilustração possibilita descrever a relação dos elementos do sistema proposto, com o intuito de trazer um sentido semântico na concepção do diagrama. A partir desse cenário, a Figura 18 foi desenvolvida para representar a interface desenvolvida nesta tese. Observa-se a existência de três atores, seis casos de uso e as associações utilizadas para determinar os responsáveis por ativar cada funcionalidade do *software*. Os atores identificados são o Solr, Usuário e Administrador, sendo que o primeiro representa uma aplicação de *software* externa e os outros dois são papéis desempenhados por indivíduos humanos. O Solr tem duas funcionalidades importantes na caracterização do *software*. Primeiramente ele é responsável pela indexação automática dos documentos a serem disponibilizados aos usuários. Esta ação é representada pelo caso de uso "Indexar documentos". Em outro momento, o Solr realiza a listagem de documentos através da ação "Listar Documentos" que corresponde à resposta para a requisição dos usuários.

Figura 18: Diagrama de caso de uso da proposta



Fonte: Elaborado pelo autor

O ator Administrador possui a função específica de estabelecer quais documentos serão indexados, selecionando a localização dos mesmos. Esta ação é determinada pelo caso de uso "Solicitar indexação". Verifica-se que ao realizar esta ação, ela está associada ao caso de uso "Indexar documentos", função realizada pelo Solr. Nessa relação existe um descritor <<include>> que é utilizado para indicar que após a requisição do administrador é obrigatória a execução da indexação provida do Solr. Também existe a possibilidade de acesso aos registros de atividades realizadas pelos usuários através da funcionalidade "Consultar registros", permitindo verificar as interações e requisições constantes no sistema.

O principal ator do sistema é denominado usuário e tem à sua disposição três casos de usos associados: "Realizar busca", "Selecionar documento" e "Acessar documento". O Quadro 6 descreve essas ações:

Quadro 6: Casos de uso do ator usuário

Função	Ação
Realizar busca	Tem a tarefa de acessar uma função advinda do Solr que faz a listagem dos documentos (Listar Documentos)
Selecionar documento	Refere-se à ação de assinalar algum documento para verificar seus atributos (características básicas) na própria tela de consulta sem a necessidade de acessar o conteúdo integral do documento
Acessar documento	Difere da anterior justamente por possibilitar acessar o documento como um todo. Esta ação pode ser realizada pelo usuário em qualquer um dos documentos listados

Fonte: Elaborado pelo autor

As ações do usuário são registradas através do caso de uso "Registrar ação" que armazena informações dos termos de busca do usuário, documentos selecionados e realmente acessados. Embora não exista identificação do usuário, cada sessão de busca é identificada e contínua durante as interações; ela só é encerrada após o usuário sair da tela de busca. Com a armazenagem das atividades dos usuários, estima-se proporcionar estudos futuros que possam analisar as interações e gerar estatísticas de busca, termos pesquisados e documentos acessados.

Conforme destacado, o diagrama de caso de uso permite verificar de maneira mais genérica quais são as funcionalidades a serem incorporadas ao *software* além, é claro, de definir os responsáveis por ativar cada uma. Este tipo de descrição se torna importante na etapa inicial do desenvolvimento, delimitando quais funções devem ser desenvolvidas e quais serão agregadas de outras aplicações. No caso desta proposta, utilizou-se o Solr como agente externo atuando como indexador de documentos e motor de busca, possibilitando que o foco seja a elaboração da interface a ser utilizada pelos usuários.

5.1.2 Diagrama de classes

O diagrama de classes pretende descrever a estrutura das informações do domínio do *software* a ser desenvolvido. Antes de detalhar a estrutura do diagrama, é fundamental esclarecer o conceito de classe para a UML. Segundo Guedes (2011), o conceito de classe surge a partir de necessidade de classificar e agrupar objetos com características e comportamentos semelhantes.

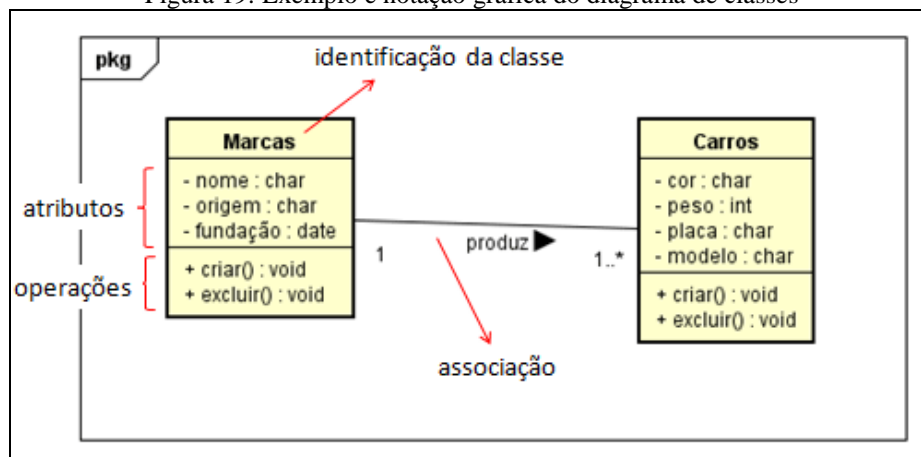
Para ilustrar este conceito utiliza-se a seguinte situação. Em um *software* destinado a controlar o aluguel de automóveis, verifica-se que os veículos a serem gerenciados possuem elementos semelhantes e, neste caso, seria denominada uma classe "Carros". Essa classe possuiria algumas características gerais como modelo, ano, cor e

número da placa, características conhecidas na UML por "Atributos". Outro conceito a ser descrito se refere às "Operações" que são os comportamentos que cada item da classe pode assumir. No caso do controle de locação de veículos, a ação de cadastrar um novo ou de alugar um veículo poderiam ser consideradas operações. Embora agrupados, os objetos dessa classe não são iguais, cada um teria sua particularidade. Um carro modelo fusca, ano 1982 de cor azul e placa AAA-9999 seria um exemplar da classe; cada exemplar da classe Carros controlado pelo *software* é nomeado como "Instância".

O diagrama de classes tem a função de apresentar uma notação gráfica das classes e as possíveis relações entre elas. Pereira (2011) salienta que existem diferentes perspectivas na concepção deste diagrama, desde apenas uma visão conceitual até uma estrutura com detalhes técnicos de implementação. Como o intuito neste estudo é o de apresentar a estrutura dos objetos e as suas relações em um nível mais abstrato e sem detalhes técnicos, será ilustrado o modelo conceitual.

Conforme descrito por Guedes (2011), o objetivo deste diagrama é determinar como as classes se complementam e trocam informações entre si, proporcionando uma visão da estrutura estática do processo. Um elemento fundamental para representar as relações são as "Associações". Elas podem determinar vínculos que ocorrem entre objetos de uma ou mais classes. Para indicar a forma de relação entre as classes, as associações podem ser rotuladas com títulos que informam o tipo de vínculo estabelecido. Podem também ser delimitadas a quantidade de objetos que participam da relação; esta indicação é nomeada como "Multiplicidade". Para esclarecer esses conceitos e a notação gráfica do diagrama foi elaborada a Figura 19 com a adição da classe "Marcas" ao exemplo dado anteriormente.

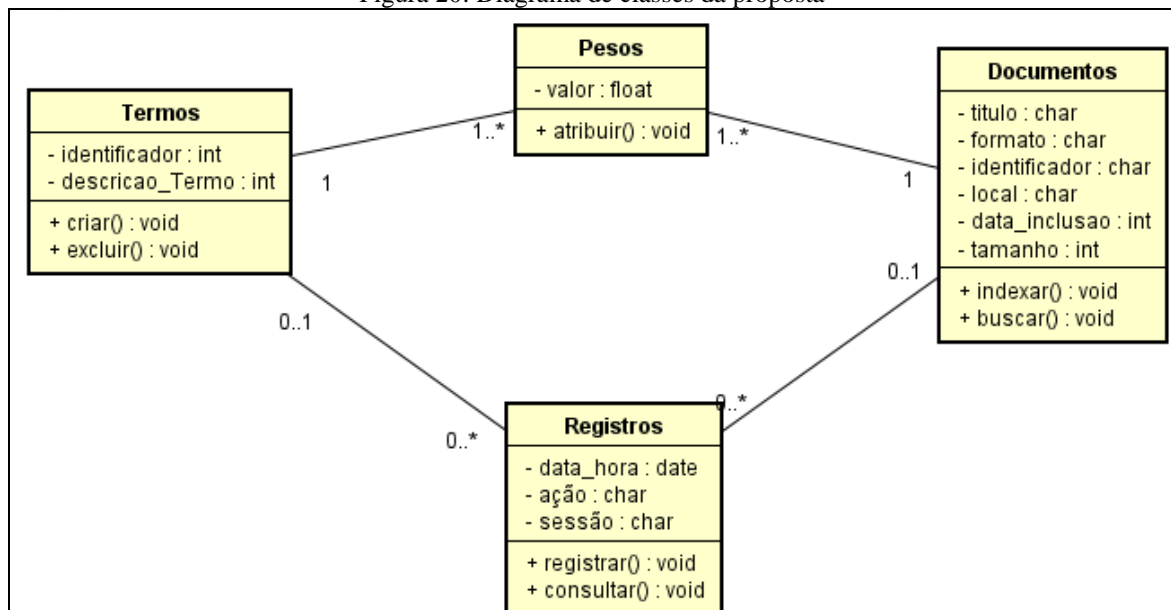
Figura 19: Exemplo e notação gráfica do diagrama de classes



Fonte: Elaborado pelo autor

Para esclarecer os requisitos e a estrutura da interface proposta nesta pesquisa, apresenta-se o diagrama de classes. Considera-se que os elementos e relações constantes nesse diagrama proporcionam desvendar a estrutura em que as informações estarão dispostas, além de estabelecer a origem dos elementos que serão exibidos na interface. Conforme já destacado, o diagrama exposto na Figura 20 apresenta as classes, atributos e operações em nível conceitual, omitindo propositalmente elementos técnicos que são utilizados na concepção do *software*, mas que não agregam significado para o entendimento da estrutura. Como exemplo dessa situação está a classe "Documentos". Aqui elenca-se apenas os atributos utilizados no processo de visualização, omitindo outros que foram extraídos no processo de indexação, mas não são utilizados nesta proposta. Os atributos apresentados são resultado do processo de indexação automática realizado pelo Solr e disponíveis para possível utilização na interface.

Figura 20: Diagrama de classes da proposta



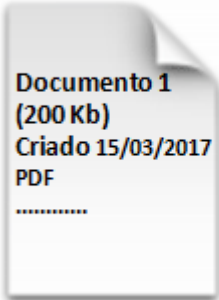
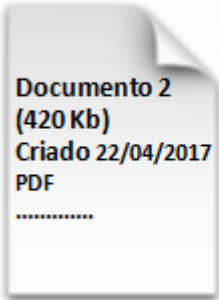
Fonte: Elaborado pelo autor

A estrutura apresentada no diagrama possui três classes (termos, pesos e documentos) que são geradas automaticamente a partir do processo de indexação realizado pelo Solr, processo que será melhor detalhado nas próximas seções. A classe registros foi gerada posteriormente com a função de armazenar dados da interação dos usuários no processo de busca, navegação e seleção dos documentos. No diagrama apresentado pela Figura 20 procurou-se abstrair e renomear esses atributos para facilitar a leitura do diagrama utilizando-se nomes em português e as devidas funções. Desta forma, aplica-se aqui a nomenclatura utilizada na programação da interface. São listados somente os atributos que

foram convertidos do formato registrado nos índices do Solr para a aplicação nesta pesquisa.

A classe documentos é utilizada para atuar na identificação dos documentos e a localização dos mesmos. Dessa forma, ao realizar uma busca, o usuário tem acesso aos dados elementares de cada documento. Observa-se que os termos que indexam o conteúdo dos documentos são tratados em outra classe denominada "Termos", a qual possibilita listar todos os termos que indexam um ou mais arquivos. A classe "Pesos" associa a lista de termos aos documentos por eles indexados; a principal ideia é fornecer a indicação de qual a importância do termo em questão para um documento específico. Para exemplificar essas associações consideram-se dois documentos (Figura 21) que hipoteticamente fazem parte de um *corpus* de artigos científicos da área de Ciência da Informação, sendo que ao final do processo de indexação poderia se obter, por exemplo, os seguintes termos de indexação e seus respectivos pesos (grau de importância do termo para o documento):

Figura 21: Exemplo de documentos indexados

 <p>Documento 1 (200 Kb) Criado 15/03/2017 PDF</p>	Termo	Peso
	Recuperação de Informação	0.85
	Peirce	0.8
	Ciência da Informação	0.7
	Interfaces	0.3
	Engenharia de Software	0.1
 <p>Documento 2 (420 Kb) Criado 22/04/2017 PDF</p>	Termo	Peso
	Interfaces	0.9
	Peirce	0.8
	Recuperação de Informação	0.3
	Visualização de Informação	0.2

Fonte: Elaborado pelo autor

Neste cenário, os dados específicos de cada documento como título, tamanho, data e formato seriam tratados pela classe "Documentos". Independentemente de constar em um ou mais documentos, os termos seriam controlados através da classe "Termos", ficando a cargo da classe "Pesos" relacionar o peso do termo ao documento específico. Verifica-se que através da notação "1..*", especificada na associação, que um

termo pode determinar pesos para vários documentos, da mesma forma que um documento tem diversos termos com pesos atribuídos a ele, o que indica o grau de importância do termo para o documento. No exemplo observa-se que cada documento é representado por diversos termos, cada qual com um grau de importância. O termo de busca "Recuperação de Informação" aparece com pesos diferentes para cada documento, sendo mais importante para identificar o "Documento 1".

A classe "Registros" não é utilizada no processo de indexação e sim durante o uso do *software*. A principal função dela é registrar as ações dos usuários a partir do início do processo de busca. Conforme já citado, os usuários não necessitam se identificar para a utilização do *software*, entretanto, ao iniciar o processo de busca, é iniciada uma sessão que é armazenada e acompanha o usuário até que ele pare sua navegação ou encerre o *software*. Através do registro de data/horário e da ação realizada, torna-se possível armazenar a sequência de pesquisa dos usuários, assim como os termos de busca utilizados e documentos selecionados e acessados.

Após apresentados os diagramas de classes e casos de uso, é possível realizar uma breve relação entre eles, conforme pode ser observado no Quadro 7.

Quadro 7: Relação entre os diagramas de casos de uso e classes

Caso de uso	Classes envolvidas
Acessar documento	Documentos, Termos, Pesos e Registros
Consultar registros	Documentos, Termos, Pesos e Registros
Indexar documentos	Documentos, Termos e Pesos
Listar documentos	Documentos, Termos e Pesos
Realizar busca	Documentos, Termos, Pesos e Registros
Registrar ação	Documentos, Termos, Pesos e Registros
Selecionar documento	Documentos, Termos, Pesos e Registros
Solicitar indexação	Documentos, Termos, Pesos

Fonte: Elaborado pelo autor

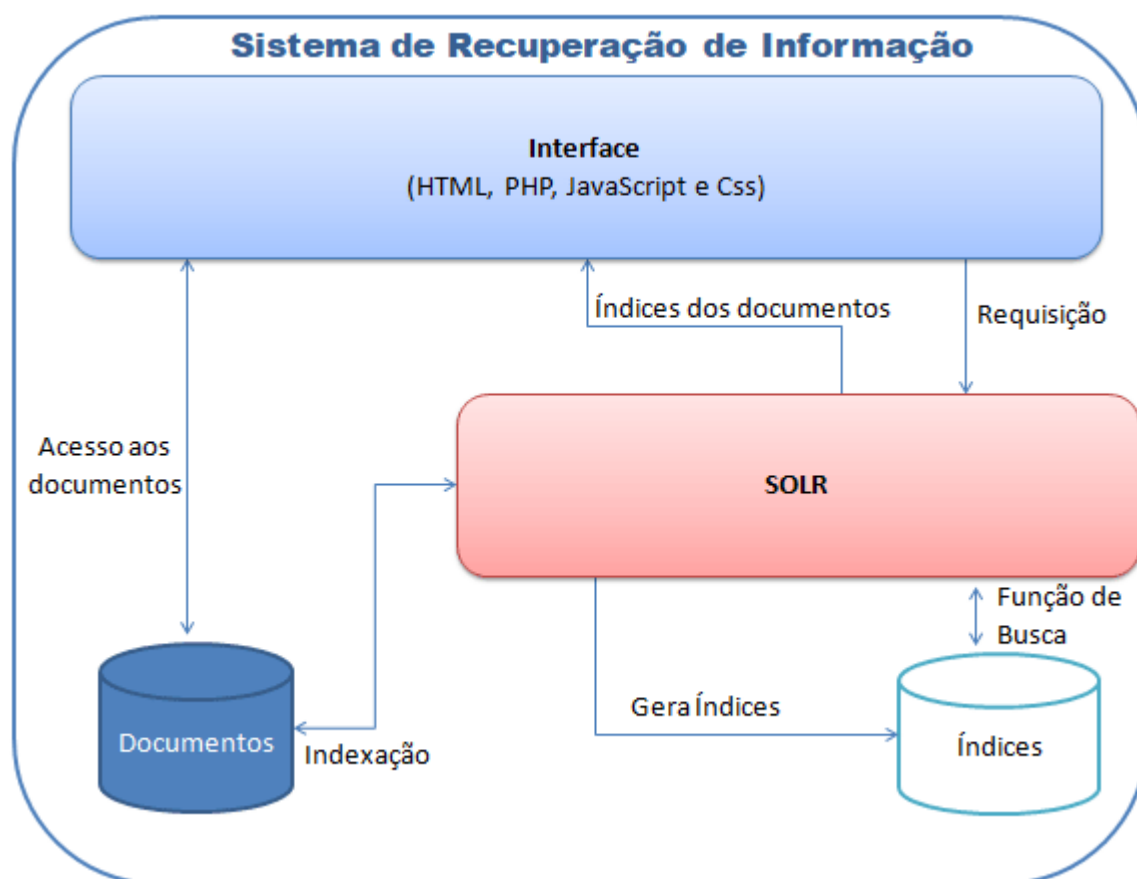
Elaborou-se o Quadro 7 com a finalidade de demonstrar que as ações realizadas e diagnosticadas no diagrama de casos de uso, tem relação direta com a utilização das classes e suas associações. A maioria das funcionalidades do *software* utiliza todas as classes para estruturar seu processamento e proporcionar o acesso às informações. Dessa forma, percebe-se a importância das associações, pois é através delas que as informações mais complexas são estruturadas, ou seja, é necessário agrupar objetos para que a informação exposta ao usuário seja completa.

Os diagramas descritos até aqui possibilitaram analisar requisitos e verificar o formato em que os dados estão conectados; a partir disso propõe-se analisar tecnologias que possibilitam transformar essa análise em uma aplicação computacional.

5.2 TECNOLOGIAS UTILIZADAS NA IMPLEMENTAÇÃO DO PROTÓTIPO

Quando se pretende desenvolver qualquer *software*, a escolha das ferramentas a serem utilizadas é um passo importante e depende dos objetivos de cada aplicação. Neste estudo, alguns fatores direcionaram a seleção das ferramentas. Inicialmente buscou-se tecnologias gratuitas, ou seja, que não demandem custos financeiros e que fornecem viabilidade técnica de serem utilizadas. Outro fator importante diz respeito à flexibilidade, visando a utilização em sistemas operacionais distintos, sem necessidade de alterações. Por fim, verificou-se quais mecanismos forneceriam suporte aos itens propostos na pesquisa, os quais são embasados nas áreas de Recuperação de Informação, Semiótica e Interação Humano Computador. A Figura 22 ilustra uma visão geral da arquitetura utilizada.

Figura 22: Arquitetura do software



Fonte: Elaborado pelo autor

Tendo em vista a utilização do modelo vetorial, foi selecionado o motor de busca Solr que proporcionou um alicerce importante no desenvolvimento da proposta. Conforme representado na Figura 22, o Solr é responsável pela função de busca e indexação dos documentos, entretanto o acesso ao documento como um todo é realizado diretamente no

local onde está armazenado. Em relação à programação e o desenvolvimento da interface, optou-se pelas tecnologias PHP (*Hypertext Preprocessor* originalmente *Personal Home Page*), HTML, JavaScript e CSS (*Cascading Style Sheets*). A interface interage diretamente com o Solr enviando requisições e recebendo os índices e dados dos documentos. A seguir são descritas e apresentadas essas tecnologias focando os aspectos que viabilizam a construção do *software* de busca.

5.2.1 Motor de busca Solr

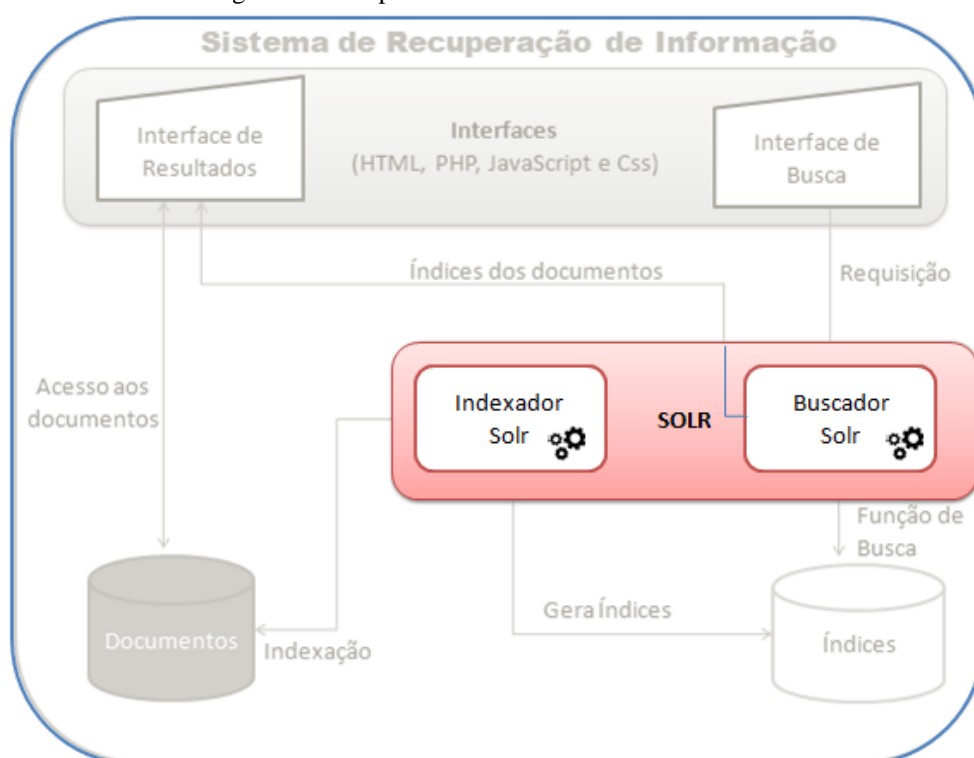
O Solr é uma plataforma de pesquisa de código aberto escrita em Java e derivada do projeto Apache Lucene. Ele permite o processamento e indexação de vários tipos de arquivos como páginas Web, documentos de texto, documentos *Microsoft Word*, documentos HTML, planilhas, documentos PDF (*Portable Document Format*), ou outros formatos que possibilitem a extração de informação textual. Outra grande vantagem do Solr é o fato de ser escalável, permitindo, por isso, que sejam efetuadas pesquisas distribuídas e que os índices estejam replicados, além disso é amplamente utilizado como padrão da indústria. Um componente central do Solr é sua biblioteca de processamento de texto extensível, que abrange uma ampla gama de funcionalidades e idiomas (SOLR, 2017).

Torres (2014) explica que a base do Solr é o projeto Apache Lucene. Isso significa que ele incorpora os princípios do Lucene e adiciona outras funcionalidades. O Lucene permite indexar e realizar pesquisa sobre os elementos indexados. Neste processo é realizada uma combinação entre o modelo booleano e o modelo vetorial. O autor também destaca que o Solr proporciona acesso aos índices através de diversos formatos como XML (*Extensible Markup Language*), Json (*JavaScript Object Notation*), PHP (*Hypertext Preprocessor*), entre outros, facilitando a reutilização de seus dados em outras ferramentas. O Solr é multiplataforma, o que proporciona ser utilizado a partir de diversos sistemas operacionais como, por exemplo, Windows e Linux. O seu código fonte é escrito na linguagem de programação Java e está disponível como um *software open source* (código aberto podendo ser adaptado para outras situações). A ferramenta também disponibiliza uma interface Web que proporciona ao administrador do sistema controlar as funções de forma mais intuitiva.

Correia (2016) pesquisou e comparou três ferramentas Solr, Elasticsearch e

SphinxSearch⁴, todas utilizadas para indexar e pesquisar conteúdos. O autor salienta que tanto o Solr como o Elasticsearch têm a licença livre, sendo que o Sphinx Search tem a licença GPLv2, que para determinados fins deve ser comprada. Mesmo tendo sua licença comercial, o Sphinx Search foi desenvolvido para atender bases de dados e, por isso, deixa a desejar na leitura de arquivos PDF, Word entre outros formatos de arquivos importantes para a pesquisa de documentos em bases com documentos diversificados. Tanto o Solr como o Elasticsearch utilizam o Apache Tika, que proporciona a extração de texto de arquivos binários, como por exemplo: PDF, Word, XML, etc. Nesse contexto, o Sphinx Search apresentou menor variedade que as outras duas ferramentas. Em termos de uso, verificou-se que o Solr e o Elasticsearch estão entre os mais utilizados, enquanto Sphinx Search tem uma popularidade bem inferior aos outros. Por fim, o autor explica que o Solr e o Elasticsearch proporcionam recursos bastante semelhantes, mas sugere a utilização do Solr por ter uma documentação mais consolidada e uma comunidade bastante ativa em seu uso. A Figura 23 destaca a estrutura do Solr.

Figura 23: Componentes Solr - Indexador e Buscador



Fonte: Elaborado pelo autor

Conforme destacado o Solr tem sua biblioteca formada por dois elementos

⁴ Solr - disponível em: <http://lucene.apache.org/solr/>
 Elasticsearch - disponível em: <https://www.elastic.co/>
 SphinxSearch - disponível em: <http://sphinxsearch.com/>

principais, ou seja, um responsável pela pesquisa e outro pela indexação. Torres (2014) descreve que o processo de indexação extrai os atributos e os dados dos documentos e utiliza-se essencialmente de palavras-chave. A pesquisa consulta o índice através dos termos requisitados pelo usuário, ordenando os resultados pela similaridade do texto com a mesma. A seguir serão detalhados os dois componentes.

5.2.1.1 Indexador

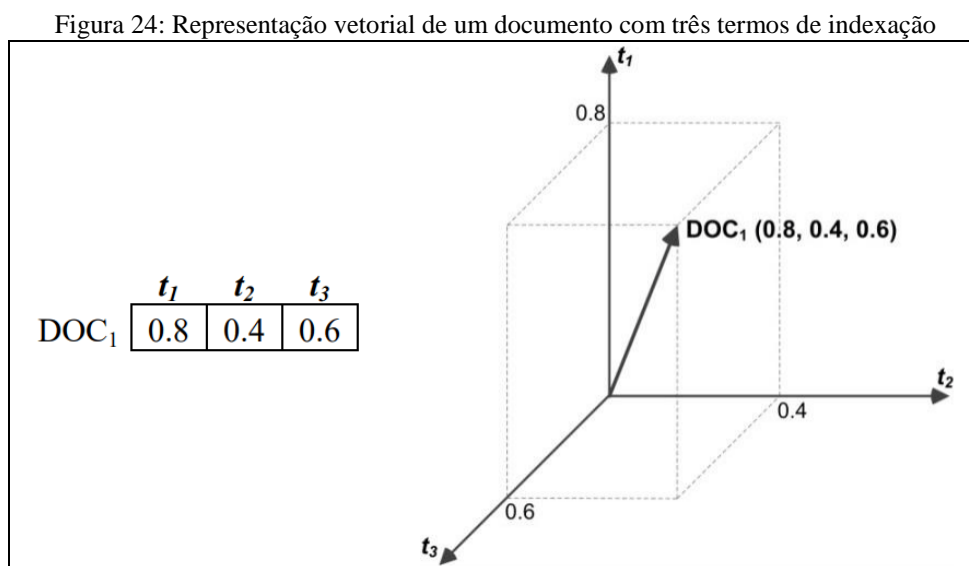
Conforme já abordado na seção **Erro! Fonte de referência não encontrada.**, o processo de indexação visa descrever um documento através das palavras-chave contidas nesses documentos. No caso do modelo vetorial, a ideia é representar através de termos extraídos do próprio documento, após sua análise. Como descrito em Lucene (2016), o Solr armazena dados em uma estrutura denominada índice invertido, que é gravado no sistema de arquivos como um conjunto de arquivos de índice. Este esquema permite aos usuários executarem procuras por palavras-chave e localizarem os documentos que correspondem a uma determinada consulta.

Antes que os termos sejam incluídos no índice, eles são processados por um analisador. Como já descrito no processo de indexação automática, na fase de análise os documentos passam por várias operações: extração das palavras, remoção de palavras comuns, redução de palavras para o formato de raiz, alteração das palavras para minúsculas. Após esta análise, os termos selecionados são incluídos no índice do Solr. Para um *corpus* heterogêneo como na Web, não é necessária a indicação formal de atributos como, por exemplo: título, nome, data de criação, tamanho do arquivo, etc., visto que por conta da diversidade os documentos não estão padronizados.

O Solr é responsável por varrer o documento em busca de atributos e também analisar o conteúdo para elencar as palavras-chave (termos) mais relevantes a cada documento. Em Solr (2017) é indicado que para realizar esta operação é utilizada uma biblioteca escrita em Java denominada "Tika," administrada pela *Apache Software Foundation* (mesma instituição que abriga o Lucene e Solr). O Apache Tika é uma estrutura de detecção e análise de conteúdo que detecta e extrai metadados e texto de mais de mil tipos diferentes de arquivos e possibilita identificar idiomas e mapear texto para campos específicos do idioma durante a indexação.

Durante o processo, além da extração dos termos que representam cada documento, é analisada a importância de cada termo para a identificação do documento.

Como já descrito na descrição do Modelo Vetorial (seção 2.2.3.2), os pesos atribuídos a cada termo seguem a ponderação TF-IDF. Relembrando este conceito, são levados em consideração a frequência com que um termo aparece em um documento, e também quanto mais raro for um termo em todos os documentos do índice. Assim, cada documento é indexado por um vetor de termos com seus respectivos pesos que serão futuramente comparados com a requisição de consulta dos usuários. Todos os elementos deste vetor assumem um valor que varia entre zero (0) e um (1). Os valores mais próximos a um sinalizam os termos mais relevantes na representação do documento. A Figura 24 ilustra a visualização em um espaço cartesiano tridimensional, a título de exemplo, três termos de indexação (t_1 , t_2 e t_3) e seus respectivos pesos 0.8, 0.4 e 0.6 que representam o DOC1 (FERNEDA, 2013, p. 79).



Embora o processo de indexação seja complexo e tenha diversas variáveis, a utilização do Solr permite realizar esta atividade através de comandos específicos. O algoritmo da ferramenta, em conjunto com a biblioteca Tika, realiza a extração dos termos e atribui os pesos que indicam a relevância de cada termo para o documento. A Figura 25 apresenta uma amostra simples de indexação de documentos constantes em um diretório.

Figura 25: Exemplo de indexação no Solr

Comando para indexação no Solr

```
java -jar -Dc=Tese -Dauto TesteSolr\post.jar TesteSolr\*.*
```

```
SimplePostTool version 5.0.0
Posting files to [base] url http://localhost:8983/solr/fabricio/update...
Entering auto mode. File endings considered are xml,json,jsonl,csv,pdf,doc,docx,
f,htm,html,txt,log
Posting file apoio.txt (text/plain) to [base]/extract
Posting file b01-Popov2004.pdf (application/pdf) to [base]/extract
Posting file b02-Kara2012.pdf (application/pdf) to [base]/extract
Posting file catalog-v001.xml (application/xml) to [base]
Posting file dubin1995.pdf (application/pdf) to [base]/extract
Posting file Fichamento.xlsx (application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) to [base]/extract
Posting file Garcia200310.1.1.436.9671.pdf (application/pdf) to [base]/extract
Posting file Marcos2005VisualInformationRetrievalSystems.docx (application/vnd.openxmlformats-officedocument.wordprocessingml.document) to [base]/extract
Posting file OliveiraJr.pdf (application/pdf) to [base]/extract
Posting file Paulheim2009-IsWC_2009.pdf (application/pdf) to [base]/extract
Posting file post.jar (application/octet-stream) to [base]/extract
Posting file PublicaçõesMID-BR.xlsx (application/vnd.openxmlformats-officedocument.spreadsheetml.sheet) to [base]/extract
Posting file Referencias.txt (text/plain) to [base]/extract
Posting file Souza2006.pdf (application/pdf) to [base]/extract
Posting file SouzaAlvarenga2004.pdf (application/pdf) to [base]/extract
Posting file Valiati2008.pdf (application/pdf) to [base]/extract
Posting file Varios2011-Ontobras.pdf (application/pdf) to [base]/extract
17 files indexed.
Committing Solr index changes to http://localhost:8983/solr/fabricio/update...
Time spent: 0:00:21.193
C:\Users\Fabricio\Documents\Fabricio\Doutorado\Solr\solr-7.2.1>
```

Arquivos indexados

Formatos TXT, PDF, XML e DOCX

Quantidade de documentos

Fonte: Elaborado pelo autor

Para este exemplo foi criado uma pasta local contendo dezessete documentos de variados formatos. Para execução desse comando é necessário especificar o local dos arquivos e também para qual coleção está sendo destinado tal índice. Especificamente nesta situação foram indexados todos os arquivos de uma pasta nomeada "TesteSolr" e adicionados ao índice da coleção "Tese".

O Solr oferece um grande nível de abstração, visto que o desenvolvedor não necessita conhecer funções e algoritmos de indexação, basta enviar e receber requisições da aplicação. A partir das considerações listadas em Solr (2017), Correia (2016), Lucene (2016) e Torres (2014), a ferramenta e sua estrutura baseada no modelo vetorial é utilizada por diversos sistemas de busca com resultados bastante satisfatórios. Os processos que utilizam cálculos matemáticos de vetores são compatíveis com as demandas computacionais atuais, além de deixar explícita a fórmula de cálculo de similaridade, deixando o processo mais transparente. Esta seção foi fundamental para conhecer os fundamentos da indexação da ferramenta e confirmar a utilização do modelo vetorial no processo, o que possibilita desvendar e justificar a ordenação dos documentos durante a busca.

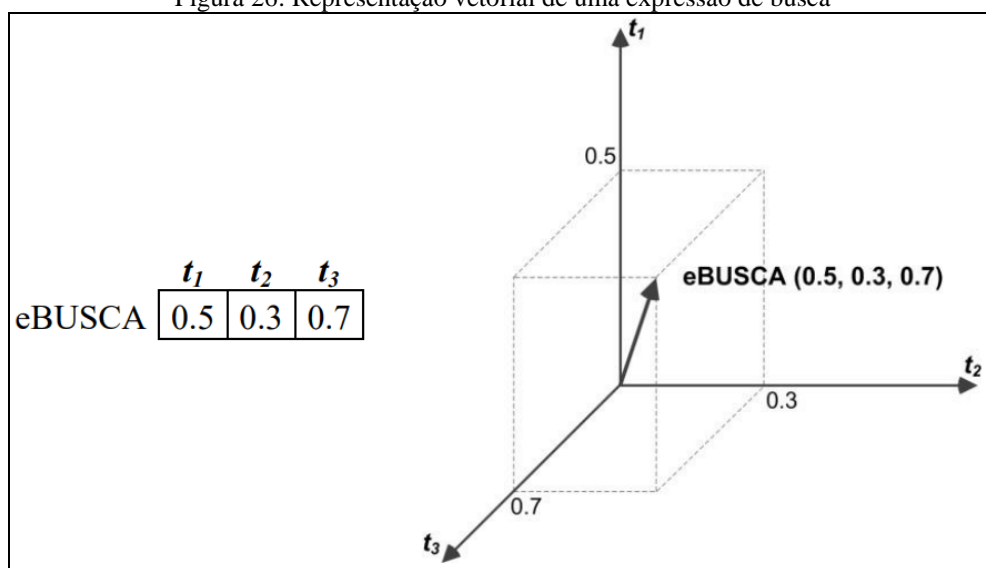
5.2.1.2 Buscador

O processo de busca considerado pelo Solr consiste em encontrar palavras no índice e relacionar os documentos que possuem essas palavras. Para isso ele oferece diferentes capacidades para expressar consultas de usuário, variando desde consultas de

palavras-chave básicas, até consultas de frases ou o uso de caracteres curingas. Para a finalidade desta pesquisa utilizaremos as buscas baseadas em palavras-chave (termos).

As pesquisas dos usuários representam suas necessidades de informação. Em geral são utilizadas palavras-chave que aspiram possuir correspondência com os documentos disponíveis, tendo diferentes graus de relevância. Em Solr (2017) é descrito que no momento da execução de uma pesquisa no Solr, a requisição é processada por um manipulador de solicitações. Este manipulador é responsável por definir a lógica a ser usada no processamento dessa solicitação. A lógica a ser utilizada baseia-se novamente no modelo vetorial, onde a requisição do usuário gera um vetor de termos que é comparado com os vetores de termos dos documentos. Dessa forma, um vetor numérico representa a expressão de busca em que cada elemento representa o grau de relevância do respectivo termo na necessidade de informação do usuário. A Figura 26 ilustra a representação vetorial de uma expressão de busca (eBUSCA) contendo três termos (t_1 , t_2 e t_3) com pesos 0.5, 0.3 e 0.7, respectivamente.

Figura 26: Representação vetorial de uma expressão de busca



Fonte: (FERNEDA, 2013, p.80)

Segundo Torres (2014), o nível de relevância é calculado com maior precisão a partir do momento que tanto consulta como documentos são representados por vetores, com isso o cálculo da distância entre os vetores é utilizado para definir a pontuação de cada documento em relação à requisição do usuário. Na geração do *score* (pontuação) dos documentos para cada requisição, é utilizada uma fórmula de similaridade (*sim*) que mede o cosseno do ângulo formado entre dois vetores, tendo como resultado um valor entre 0 e 1. Ferneda (2013, p.81) destaca que "quanto menor for o ângulo, maior será a similaridade entre

eles". Em resumo, a similaridade entre o documento DOC1 e a expressão eBUSCA, ilustrados respectivamente nas Figuras (Figura 24 e Figura 27), é calculada como:

$$\text{DOC}_1 \begin{array}{|c|c|c|} \hline t_1 & t_2 & t_3 \\ \hline 0.8 & 0.4 & 0.6 \\ \hline \end{array} \quad \text{eBUSCA}_1 \begin{array}{|c|c|c|} \hline t_1 & t_2 & t_3 \\ \hline 0.5 & 0.3 & 0.7 \\ \hline \end{array}$$

$$\text{sim}(\text{DOC}_1, \text{eBUSCA}) = \frac{(0.8 \times 0.5) + (0.4 \times 0.3) + (0.6 \times 0.7)}{\sqrt{(0.8^2 + 0.4^2 + 0.6^2)} \times \sqrt{(0.5^2 + 0.3^2 + 0.7^2)}} \cong 0.96$$

A título de exemplo, podemos recorrer à situação apresentada na Figura 21 e adaptada para a Figura 27 que exemplifica "Documento 1" e "Documento 2" indexados por diversos termos e seus respectivos pesos. Supondo que se realize uma pesquisa utilizando o termo "Recuperação de Informação", nesta situação o Solr recebe a requisição e realiza o cálculo de similaridade entre o vetor que representa a consulta e os documentos do corpus.

Figura 27: Exemplo de busca no Solr

Busca pelo termo "Recuperação de Informação"			Score (Pontuação)
Documento 1 (200 Kb) Criado 15/03/2017 PDF	Termo	Peso	→ 0.935
	Recuperação de Informação	0.85	
	Peirce	0.8	
	Ciência da Informação	0.7	
	Interfaces	0.3	
	Engenharia de Software	0.1	
Documento 2 (420 Kb) Criado 22/04/2017 PDF	Termo	Peso	→ 0.42
	Interfaces	0.9	
	Peirce	0.8	
	Recuperação de Informação	0.3	
	Visualização de Informação	0.2	

Fonte: Elaborado pelo autor

Nota-se que a expressão informada pelo usuário "Recuperação de Informação" tem um maior grau de importância na representação do "Documento 1", então seria gerado uma pontuação de ranqueamento mais elevada para ele, em relação ao "Documento 2". Como na indexação, o processo de busca é viabilizado pelo Solr sem a necessidade de realizar qualquer tipo de programação ou de conhecimento do algoritmo de ranqueamento, é necessário enviar a requisição e a aplicação retorna os atributos dos documentos selecionados a partir do índice.

Figura 28: Exemplos de resposta do Solr

Resposta do Solr - Busca pelo termo "Recuperação de Informação"	
Formato XML	Formato PHP
<pre><lst name="response"> <bool name="zkConnected">true</bool> <int name="status">0</int> <int name="QTime">148</int> <lst name="params"> <str name="q">Recuperação de Informação</str> <str name="wt">xml</str> <str name="_">1531420640020</str> </lst> </lst> <result name="response" numFound="45" start="0"> <doc> <str name="id">C:\Users\Fabricio\Documents\Docume</pre>	<pre>array('responseHeader'=>array('zkConnected'=>true, 'status'=>0, 'QTime'=>602, 'params'=>array('q'=>'Recuperação de Informação', 'wt'=>'php', '_'=>'1531420640020'), 'response'=>array('numFound'=>45,'start'=>0,'maxS array('id'=>'C:\\Users\\Fabricio\\Documents\\Fabr 'date'=>array('2010-03-16T18:05:50Z'),</pre>

Fonte: Elaborado pelo autor

Em uma comparação rápida dos dois formatos, constata-se que o XML retorna os dados utilizando etiquetas (*tags*) delimitadas pelos caracteres "< >". No caso do formato em PHP, o retorno é um vetor (*array*). Seja um vetor ou através de etiquetas, a disposição dos elementos permite a manipulação dos dados para os fins desejados e, por consequência, arranjá-los na interface para possibilitar formas diferentes de visualização. Importante salientar que além da relação de documentos, são apresentados dados referentes à consulta como status, o tempo de consulta, quantidade de documentos, etc. Conforme descrito por Solr (2017), os vários formatos de resposta têm a finalidade de atender diferentes tipos de aplicações, sejam para web ou sistemas locais. Dessa forma, existe uma modularização, ou seja, uma independência dos dados em relação à interface. A próxima seção apresenta linguagens utilizadas para prover acesso e estruturar a visualização dos resultados.

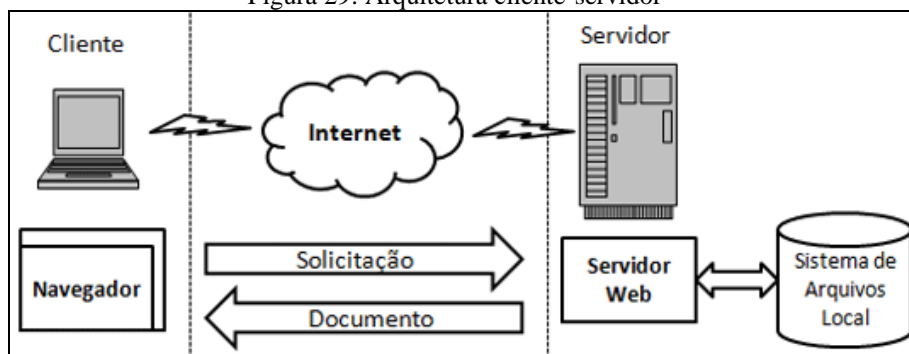
5.2.2 Tecnologias para interfaces Web

Diante de diversas tecnologias dispostas para o desenvolvimento de *softwares*, apresenta-se nesse item uma descrição das ferramentas selecionadas para este projeto. Para Chaubey (2001), as aplicações desenvolvidas para Web estão cada vez mais difundidas em comparação com os aplicativos tradicionais. Isso ocorre pelo reconhecimento das vantagens que a Internet oferece. O autor explica que a web permitiu fundamentalmente novas possibilidades para criar, impulsionar e alcançar diferenciação no domínio do negócio. Outro fator, é a infinidade de ferramentas e tecnologias que podem ser aproveitadas para o rápido desenvolvimento e implantação de aplicativos. Um aspecto importante no desenvolvimento baseado na web é o requisito de integração de componentes prontos para uso

e serviços adquiridos com outros módulos funcionais do aplicativo.

O item 5.2.1 anterior descreveu o motor de busca Solr que serve como alicerce para a aplicação de recuperação de informação. Torna-se necessário utilizar outras tecnologias que possibilitem a visualização adequada dos resultados de busca realizados na ferramenta. Primeiramente é válido esclarecer que será utilizada uma arquitetura denominada cliente-servidor (Figura 29). Millete e Bertagnolli (2014) explicam que as aplicações web atuam através de requisições de um cliente ao servidor que recebe a requisição e envia uma resposta ao cliente. Esta resposta pode ser desde uma página HTML, até uma página de erro. Os usuários atuam ao lado do cliente através de navegadores de internet.

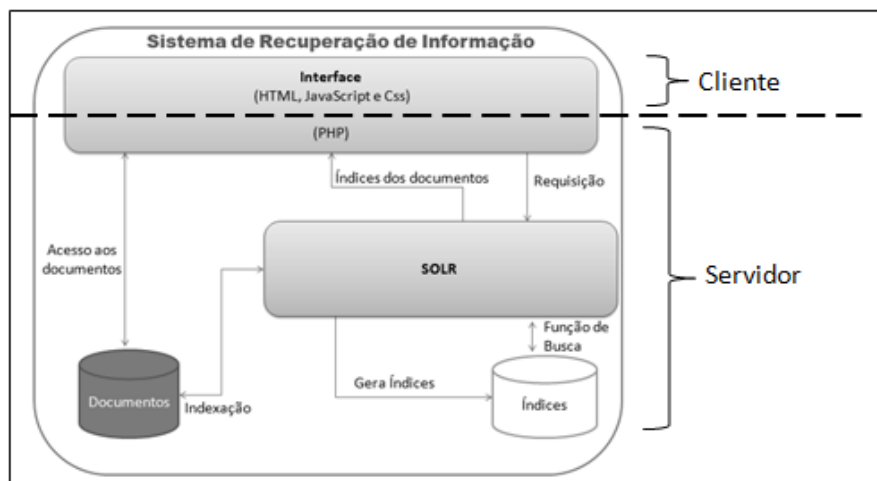
Figura 29: Arquitetura cliente-servidor



Fonte: Millete e Bertagnolli (2014)

Compreender esta estrutura é fundamental para desenvolver aplicações Web, pois é necessário combinar elementos que atuam tanto no lado do cliente como no lado do servidor. No caso desta pesquisa, o Solr é configurado no servidor (Figura 30) onde ficam armazenados os índices e os documentos a serem acessados.

Figura 30: Arquitetura cliente-servidor da proposta



Fonte: Elaborado pelo autor

A Figura 30 demonstra como o *software* desta pesquisa está estruturado em relação à arquitetura cliente-servidor. Os documentos, índices e o Solr estão armazenados no servidor, além disso a elaboração da interface necessita de especificações que também são executadas no servidor; neste caso utiliza-se o PHP. No cliente as instruções são realizadas através de páginas escritas em HTML com auxílio do CSS e JavaScript. Tais tecnologias serão detalhadas a seguir.

5.2.2.1 HTML, CSS e JavaScript

O desenvolvimento de páginas para Web pode ser dividido em três camadas básicas: a estrutura dos dados disposto pelo HTML, a formatação definida pelo CSS e o comportamento dos elementos atribuído ao JavaScript. No capítulo 3 foi fundamentada a IHC, e um dos fatos históricos destacados na concepção da área foi a criação do HTML (*hyper text mark up language*), em 1989, por Tim Berners-Lee. Durante essas décadas houve grande evolução nessa linguagem, entretanto os princípios de conexão entre páginas se mantiveram.

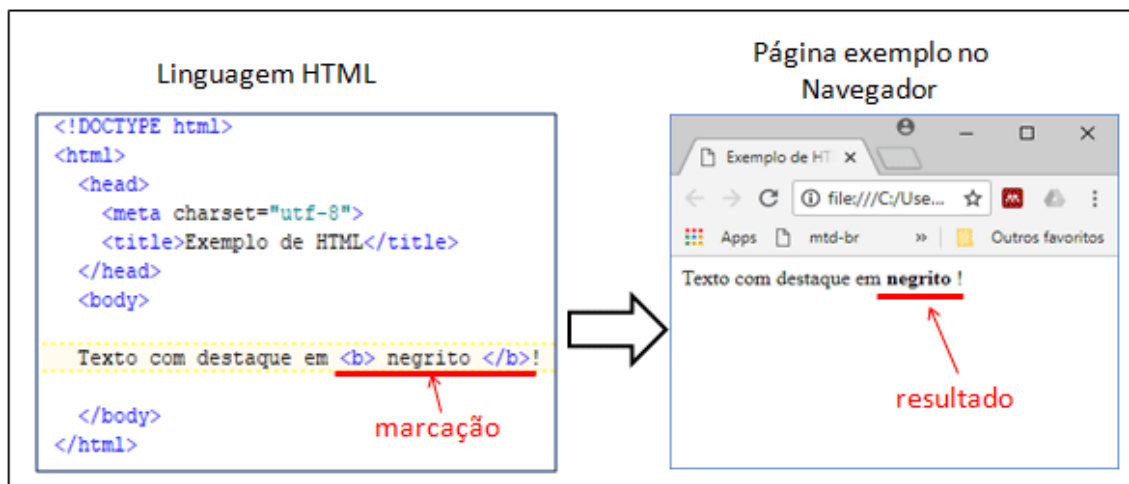
Milleto e Bertagnolli (2014) esclarecem que o HTML é a base para a elaboração de páginas Web que são exibidas pelos navegadores de internet. É conhecida como uma linguagem de marcação composta por etiquetas (*Tags*) que possibilitam a inserção de recursos como imagens, links, tabelas, textos e vídeos, porém a sua apresentação visual é restrita. Utiliza-se então o CSS e o JavaScript para agregar elementos de formatação e interação às páginas, tornando mais interativo o acesso por parte dos usuários. Como destacado na seção anterior (Figura 30), o HTML, CSS e o JavaScript operam ao lado do cliente, ou seja, são interpretados pelos navegadores utilizados pelos usuários.

A linguagem HTML é formada por um conjunto de *tags* responsável pela marcação do conteúdo de uma página Web. As *tags* são elementos interpretados pelos navegadores que exibem o conteúdo de acordo com a funcionalidade designada, seja para destacar uma palavra ou inserir uma imagem, por exemplo. A sintaxe de uma *tag* é demarcada utilizando os caracteres "<" e ">" contendo a função entre esses caracteres. Embora não seja objetivo apresentar uma visão profunda das especificações HTML, é importante exemplificar a sua sintaxe para possibilitar uma compreensão básica desta linguagem.

A Figura 31 representa como as *tags* são interpretadas pelos navegadores. O código descrito em HTML apresenta alguns elementos iniciais indicando que se trata de um documento HTML, descrevendo o título da página, em seguida um simples texto contendo a *tag* "" para apontar o trecho que será destacado em negrito. Nota-se que ao indicar a *tag*

"", encerra-se a formatação. Como é uma linguagem padrão para as páginas *Web*, esta estrutura independe de sistema operacional ou da utilização de um navegador específico. Assim, a formatação é a mesma, seja qual for a combinação de *softwares* utilizados no cliente.

Figura 31: Exemplo HTML



Fonte: Elaborado pelo autor

Inicialmente a linguagem HTML atuava na estruturação e também formatação visual das páginas, entretanto descrever a organização visual mesclada com conteúdo de texto se mostrou algo bastante trabalhoso. Segundo Milleto e Bertagnolli (2014), o *Cascading Style Sheets* (CSS) surgiu com o objetivo de cuidar da estilização da página e, nesta tarefa, o CSS é mais robusto que o HTML. Com isso as regras de estilo não aparecem mais no HTML, apenas no CSS, o que proporciona maior flexibilidade e controle na especificação. Outra vantagem é o compartilhamento do formato dentro de um mesmo projeto. Como o CSS pode ser escrito em um arquivo separado da estrutura, é possível reutilizar as mesmas especificações em diversas páginas *Web* diferentes. A Figura 32 exemplifica essa situação:

Figura 32: Exemplo de estilo utilizando CSS

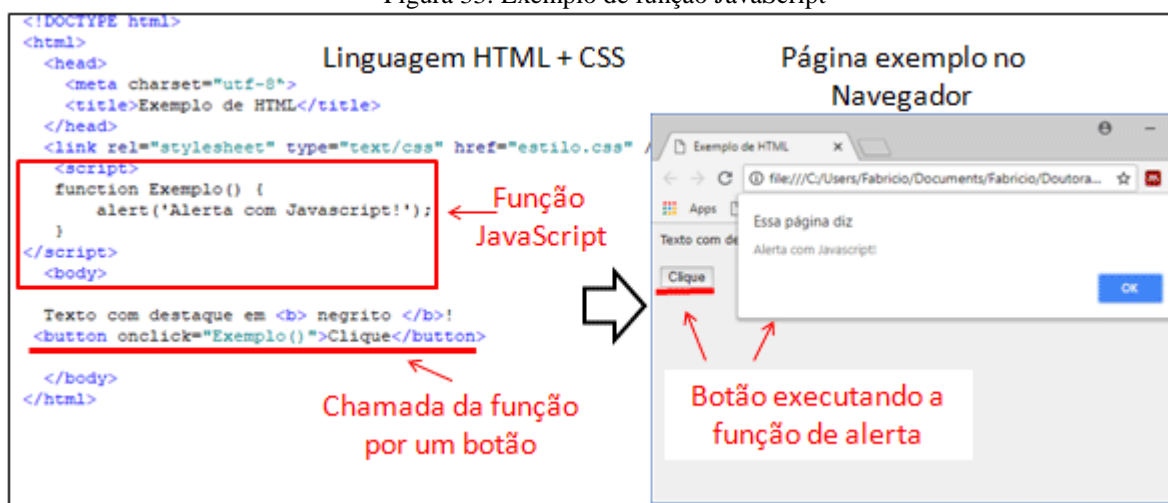


Fonte: Elaborado pelo autor

Utilizou-se o mesmo exemplo da Figura 31, incorporando o comando em destaque que realiza a chamada do arquivo de estilos denominado "estilos.css". A Figura 32 apresenta os códigos do CSS que realizam três modificações: cor de fundo da página, estilo e tamanho da fonte que formata o texto. A cada página construída, basta realizar a chamada ao arquivo "estilo.css" que a mesma passa a utilizar o formato. Ao mesmo tempo, todas as alterações realizadas no estilo serão replicadas a todas as páginas que o utilizam. Obviamente estão descritos poucos comandos, mas o CSS pode controlar fontes, cores, margens, linhas, alturas, larguras, imagens de fundo, posicionamentos, entre outras situações.

Na terceira camada de desenvolvimento está o componente JavaScript que inicialmente foi concebido para proporcionar uma interação maior do usuário com as páginas. A W3schools (2018) a caracteriza como uma linguagem considerada *client-side*, ou seja, realiza funções interpretadas no computador-cliente pelos navegadores. Ela é utilizada para controlar o HTML e o CSS, agindo na manipulação de comportamentos na página. No JavaScript o código é interpretado e executado à medida que é reconhecido pelo navegador, linha a linha, assim como o HTML.

Figura 33: Exemplo de função JavaScript



Fonte: Elaborado pelo autor

A Figura 33 ilustra a adição de um botão associando a função de exibir uma alerta na página. No exemplo, o código JavaScript está incorporado à página através das marcações "`<script>`" e "`</script>`", entretanto da mesma forma que o estilo CSS, os códigos podem ser desenvolvidos em outro arquivo, o que possibilita também o reuso das funções criadas. Millete e Bertagnolli (2014) explicam que além de executar funções, essa linguagem também permite a criação de conteúdo que se atualiza dinamicamente, controlar multimídias, imagens animadas, operações matemáticas, transformações de textos e determinar resposta a eventos que ocorrem em uma página da Web. Essas funcionalidades promovem uma variedade de opções para prover interatividade às páginas.

Conclui-se que a combinação das três tecnologias, HTML, CSS e JavaScript atendem aos objetivos almejados na interface no que diz respeito ao desenvolvimento da interface a ser interpretada nos navegadores dos usuários. Por fim, ainda existe uma camada desenvolvida no lado do servidor que efetua as requisições ao Solr e converte suas respostas nas páginas enviadas ao cliente. Esta função será executada pela linguagem de programação PHP que será caracterizada no tópico 5.2.2.2.

5.2.2.2 PHP

A estrutura apresentada pela Figura 30 ilustrou as tecnologias aplicadas na construção do *software* proposto. Evidenciando em especial o segmento relativo à interface, o PHP é o único elemento executado no servidor da aplicação. Niederauer (2011) destaca que o PHP é uma das linguagens de programação mais utilizadas na Web. Esta linguagem possibilita o desenvolvimento de variadas aplicações, a exemplo de websites dinâmicos,

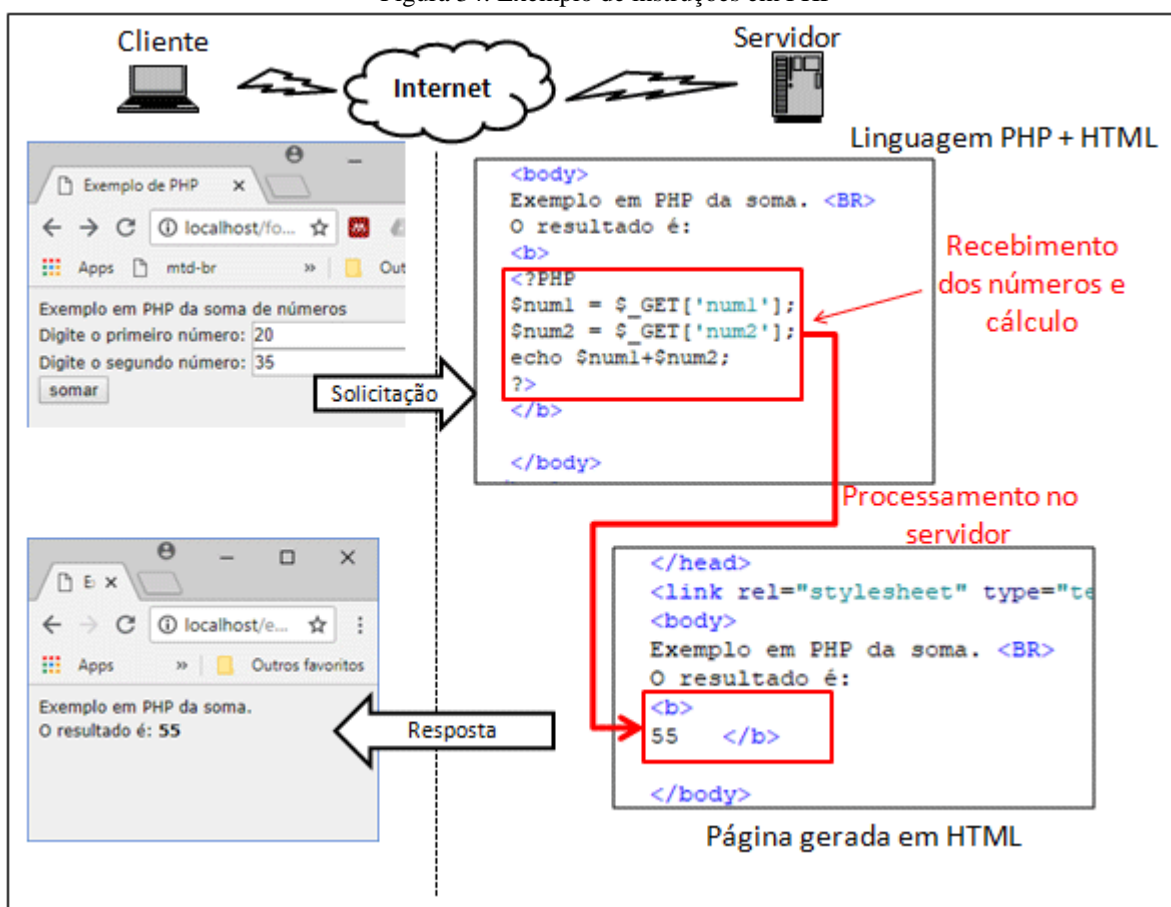
permitindo a interação por meio de links, formulários e parâmetros enviados pelo usuário. Os comandos escritos são armazenados em computadores servidores que interpretam as instruções de acordo com as requisições dos usuários e respondem com páginas em formato HTML, lidas pelos navegadores de internet disponíveis aos usuários.

Dessa forma é possível interagir com aplicações instaladas no servidor como bancos de dados, programas de e-mail, gerenciamento de arquivos e, no caso desta pesquisa, o Solr. Especificamente nesta proposta, o PHP tem a função de receber os termos de busca definidos pelos usuários e interagir diretamente com o Solr, recebendo como resultado a lista de documentos e seus atributos, gerando a interface de resultado.

Embora escrita no lado do servidor, esta linguagem é desenvolvida em conjunto com o código HTML. Para delimitar os segmentos do código PHP são utilizadas as *tags* "<?php" e "?>" que indicam início e fim das instruções. Diferentemente do HTML, CSS e JavaScript que podem ser acessadas e visualizadas pelos usuários nos navegadores, uma instrução escrita em PHP não trafega pela internet. Ao receber uma requisição o *software* PHP, localizado no servidor, executa as instruções gerando dinamicamente trechos em HTML, de acordo com a operação requisitada.

O processo estruturado na Figura 34 apresenta uma solicitação de um operação matemática simples. O usuário indica dois números e faz uma solicitação através do botão "somar". Neste momento o navegador envia esta requisição ao servidor. No exemplo, o documento em PHP tem instruções para receber os números, efetuar a soma e gerar a resposta que, neste caso, é apresentada através do comando "echo". Após gerada a resposta, a página é enviada ao navegador do usuário com o soma calculada e em formato HTML. Desta forma pode ser apresentada ao usuário utilizando as formatações cabíveis. Obviamente este tipo de operação poderia ser perfeitamente executada no próprio navegador através de código JavaScript, entretanto alguns processos mais complexos que dependem de dados do servidor, necessitam de uma linguagem de programação como o PHP que realize este processamento tornando-o transparente para os usuários finais .

Figura 34: Exemplo de instruções em PHP



Fonte: Elaborado pelo autor

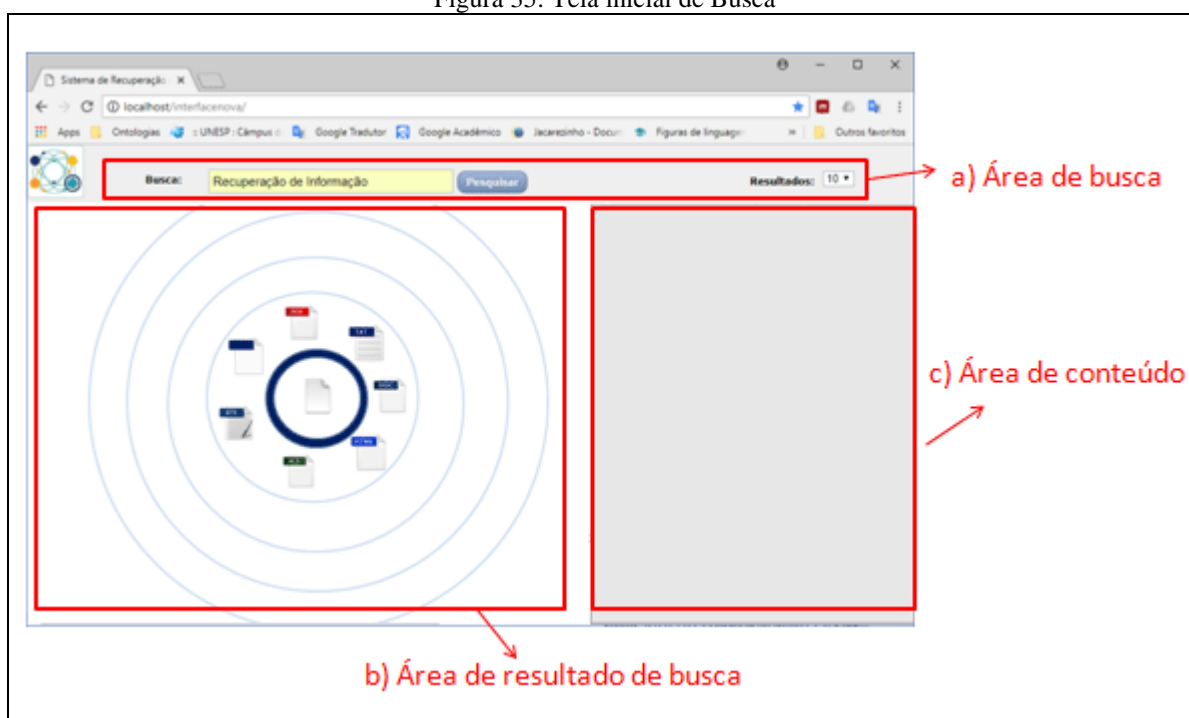
Através dessas tecnologias, o acesso ao sistema pode ser feito a partir dos navegadores de internet mais utilizados no mercado como Google Chrome, Safari Firefox e Internet Explorer, por exemplo. Não existe a necessidade de instalação de qualquer *software* adicional nos clientes, independente também do sistema operacional utilizado. Conforme será apresentado na próxima seção, o PHP possui as funcionalidades necessárias para integração ao Solr, assim como o HTML, o CSS e o JavaScript se conectam para possibilitar aos usuários meios de visualização e interação com os sistemas *Web*.

5.3 PROTÓTIPO

Esta seção tem o intuito de apresentar a interface proposta através de imagens e códigos desenvolvidos na elaboração. Além da interface em si, pretende-se demonstrar a interligação das tecnologias utilizadas no desenvolvimento, apresentando formatos e fluxos de dados existentes na arquitetura do *software*. Neste momento não serão abordados os aspectos teóricos da RI, IHC e semiótica que embasaram a utilização de

elementos visuais e de interação dispostos na interface; este suporte será discutido na próxima seção. Atenta-se então aos elementos mais técnicos, buscando ilustrar a combinação do HTML, CSS, JavaScript e PHP, e também a comunicação com o motor de busca e indexação Solr.

Figura 35: Tela inicial de Busca



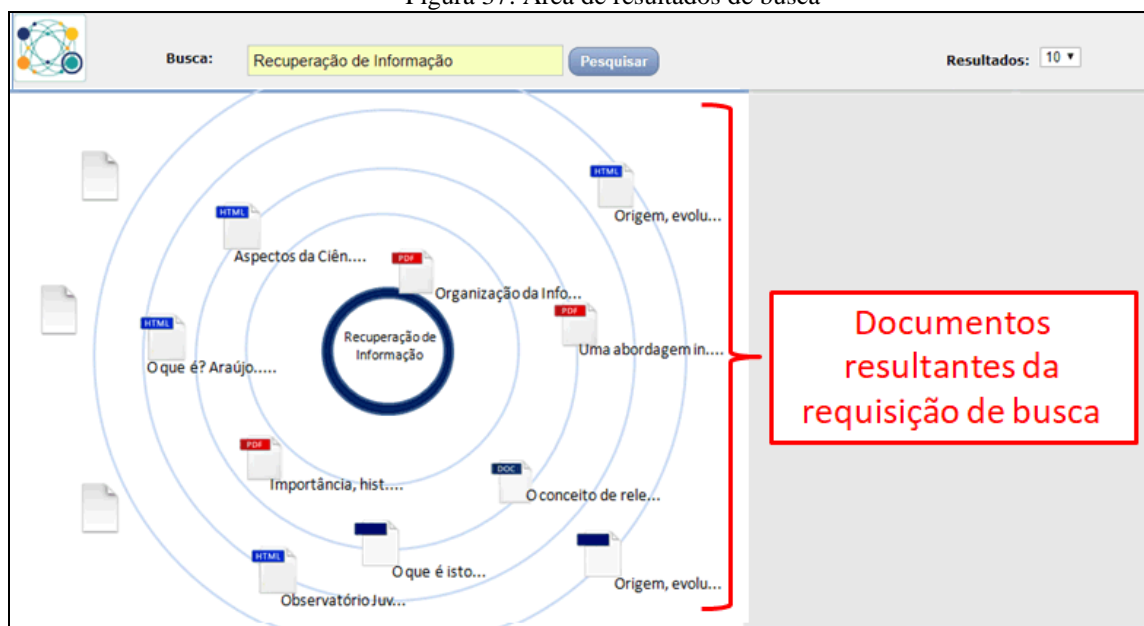
Fonte: Elaborado pelo autor

A interface proposta foi implementada sem o uso de codificações prontas de interface, pois havia a necessidade de adicionar as especificidades do Solr e acrescentar as características próprias da pesquisa. Inicia-se esta apresentação através da Figura 35, composta por três áreas que serão utilizadas como base para todo o processo de busca e interação: a) área de busca, b) área de resultado de busca e c) área de conteúdo.

A partir da intenção do estudo específico da área de resultados, utiliza-se na área de busca o conceito do retângulo de busca apresentado por Baeza-Yates e Ribeiro Neto (2013) que se tornou popular em função da sua simplicidade. Neste caso é apresentada uma caixa de texto onde o usuário especifica suas necessidades através de palavras ou termos compostos, utilizando a linguagem natural. Conforme apresentado anteriormente como um dos estilos de interação, a linguagem natural é bastante atrativa para os usuários pela simplicidade em sua utilização. Por conta desta simplicidade, é natural que existam discussões sobre este tipo de interação; aqui ela é utilizada para compor e tornar o *software* funcional, assim, não será discutida sua eficiência. Entretanto, é importante salientar que conforme

determinar que essa caixa de texto será estilizada a partir das instruções constantes na categoria "input" do arquivo CCS. Essa categoria determina alguns fatores como a fonte sendo do tipo "Arial", o tamanho da fonte com 15 pixels entre outras definições. De certa forma, a codificação dessa região da página é bastante simples, pois inicialmente não existe processamento, ficando a cargo do navegador apenas interpretar e apresentar os elementos. As requisições enviadas pelo usuário são processadas e apresentadas na área de resultados de busca.

Figura 37: Área de resultados de busca



Fonte: Elaborado pelo autor

Em destaque na Figura 37 está a área de resultados. Após enviada a requisição do usuário para o servidor, ela é processada e a resposta é apresentada através de documentos que orbitam o termo de busca disposto na área central. O processamento para a apresentação dos resultados é realizado no servidor através da linguagem PHP, combinada ao motor Solr. É executada uma instrução PHP ao Solr que retorna uma matriz contendo os documentos e atributos como nome, localização, tamanho, formato, termos que representam o documento, entre outras características. Conforme será apresentado oportunamente, a disposição dos documentos difere dos modelos tradicionais no sentido de que esses atributos são apresentados no momento em que o usuário repousa o mouse sobre o documento de interesse. Embora existam diversas instruções necessárias, a Figura 38 destaca o trecho do código desenvolvido em PHP que realiza a requisição e recebe a resposta do Solr.

Figura 38: Código PHP

Código PHP

```

<?php
include("funcoes.php");
$termo = removeAcentos($_GET["termo"]);
$linhas = $_GET["rows"];

$colecao = 'testeSolr';

$urlxml = "http://localhost:7574/solr/".$colecao."/select?q=".str_replace('
$matriz = file_get_contents($urlxml);

eval("\$result = " . $matriz . "");
echo "Total de documentos Encontrados: ".$result["response"]["numFound"]."<BR>";
$maxscore = $result["response"]["maxScore"];
echo "Maior Score: ".$maxscore."<BR><HR>";

if ($result["response"]["numFound"]<$linhas)
{
    $qtde=$result["response"]["numFound"];
}
else
{
    $qtde = $linhas;
}

```

Requisição ao Solr e recepção do resultado em uma matriz de documentos e seus atributos.

Fonte: Elaborado pelo autor

Como pode ser observado na primeira linha em destaque, o PHP faz uma requisição utilizando o protocolo de internet HTTP (*HyperText Transfer Protocol*), em tradução para o português, "Protocolo de Transferência de Hipertexto". O HTTP é o padrão utilizado na internet para transferência de dados. Após realizar a requisição, o Solr responde com uma matriz de resultados que pode ser observada de forma sintética na Figura 39.

Figura 39: Matriz de resultados do Solr

```

array(
  'responseHeader'=>array(
    'zkConnected'=>true,
    'status'=>0,
    'QTime'=>57,
    'params'=>array(
      'q'=>'recuperacao de informacao',
      'fl'=>'score, id, content_type,xmptpg_npages',
      'Submit'=>'Pesquisar',
      'rows'=>'10',
      'wt'=>'php'),
  'response'=>array('numFound'=>49, 'start'=>0, 'maxScore'=>5.0308957, 'docs'=>array(
    array(
      'id'=>'C:\Users\Fabricio\Documents\Fabricio\Doutorado\Solr\solr-7.2.1\testeSolr\apoio.txt',
      'content_type'=>array('text/plain; charset=windows-1252'),
      'score'=>5.0308957),
    array(
      'id'=>'C:\Users\Fabricio\Documents\Fabricio\Doutorado\Solr\solr-7.2.1\testeSolr\Carvalho2010.pdf',
      'content_type'=>array('application/pdf'),
      'xmptpg_npages'=>array(16),
      'score'=>3.4827342),
    array(
      'id'=>'C:\Users\Fabricio\Documents\Fabricio\Doutorado\Solr\solr-7.2.1\testeSolr\OliveiraJr.pdf',
      'content_type'=>array('application/pdf'),
      'xmptpg_npages'=>array(17),
      'score'=>3.1138384),
    array(
      'id'=>'C:\Users\Fabricio\Documents\Fabricio\Doutorado\Solr\solr-7.2.1\testeSolr\OliveiraJr.pdf',
      'content_type'=>array('application/pdf'),
      'xmptpg_npages'=>array(19),
      'score'=>3.057751),
    array(

```

Dados gerais da consulta

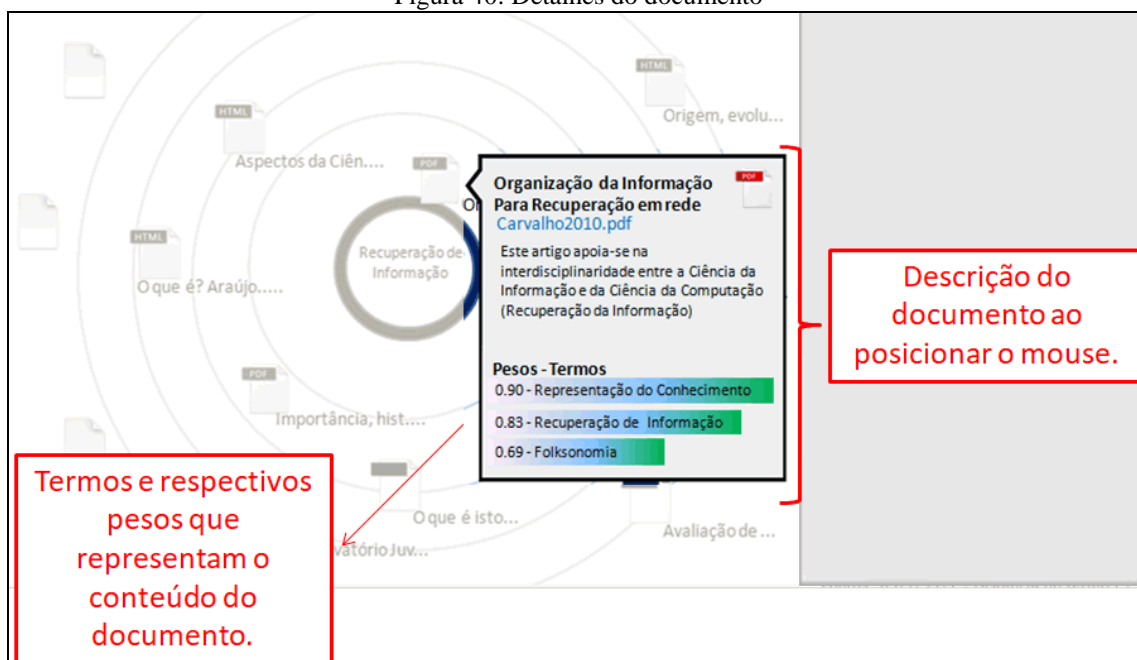
Matriz com os documentos resultantes

Fonte: Elaborado pelo autor

A situação ilustrada é a busca pelo termo "Recuperação de Informação". No cabeçalho da resposta constam dados gerais como tempo de resposta, quantidade de documentos, maior pontuação (*score*) de documento. Além de seus atributos, os documentos retornados possuem uma pontuação e quanto maior essa pontuação, considera-se que o documento tem maior relevância em relação à requisição do usuário. Dessa forma, a maior pontuação é posicionada mais próxima ao centro, enquanto a menor fica mais distante. Esta pontuação é calculada pelo Solr através do modelo vetorial; já a disposição dos elementos é realizada com base na técnica de visualização de informação denominada RadViz.

Após a listagem dos documentos, o usuário tem as opções de selecionar ou acessar um documento. No caso da seleção, basta posicionar o cursor do mouse sobre o documento que lhe serão apresentados detalhes como nome, uma breve descrição, os termos que indexam cada documento e qual o peso de cada termo. Esses detalhes podem ser observados na Figura 40. As informações exibidas são extraídas da matriz de dados que o Solr gera para cada documento e são dispostas utilizando as linguagens já nominadas para implementação da interface.

Figura 40: Detalhes do documento



Fonte: Elaborado pelo autor

Já o acesso ao documento é realizado após clicar no ícone que representa o mesmo. Neste caso o documento é aberto por padrão na região de conteúdo disposta ao lado direito do navegador. O usuário pode utilizar também as opções que os navegadores de internet proporcionam para abrir o documento em um uma nova aba ou outra janela. A

exibição do documento, na íntegra, está caracterizado na Figura 41.

Figura 41: Acesso ao documento completo

Busca: Resultados: 10

Organização da Informação Para Recuperação em rede
Carvalho2010.pdf
Este artigo apoia-se na interdisciplinaridade entre a Ciência da Informação e da Ciência da Computação (Recuperação da Informação)

Pesos - Termos
0.90 - Representação do Conhecimento
0.83 - Recuperação de Informação
0.69 - Folksonomia

Palavras-chave: Representação do Conhecimento; Recuperação da Informação; Folksonomia.

INTRODUÇÃO

A organização do conhecimento e a recuperação da informação passam pela história da classificação do conhecimento humano, desde o princípio da institucionalização da ciência e a necessidade de classificá-la visando sua posterior recuperação. Com a expansão da produção do conhecimento surge a necessidade da criação de métodos, técnicas e sistemas que possibilitem a recuperação de informações.

Os sistemas de organização do conhecimento evoluíram consideravelmente desde o surgimento da internet. No contexto da World Wide Web (web) surgiram por exemplo, as redes *peer-to-peer*, cuja arquitetura caracteriza-se pela descentralização das

funções na rede e permite a classificação do conteúdo pelos pares que produzem e compartilham informações. Neste ambiente de compartilhamento a intermediação de atores que

Documento apresentado na íntegra ao ser selecionado.

Fonte: Elaborado pelo autor

Por fim, existe a opção de um atalho para pesquisa. Ao verificar as palavras que indexam o documento, pode-se clicar sobre elas abrindo uma nova pesquisa com o termo escolhido.

Figura 42: Pesquisa a partir de palavra-chave

Organização da Informação Para Recuperação em rede
Carvalho2010.pdf
Este artigo apoia-se na interdisciplinaridade entre a Ciência da Informação e da Ciência da Computação (Recuperação da Informação)

Pesos - Termos
0.90 - Representação do Conhecimento
0.83 - Recuperação de Informação
0.69 - Folksonomia

A selecionar qualquer termo pode ser realizada nova busca.

Recuperação...

Folksonomia

Definição de ...

História...

O conceito de...

Relações da Folk...

Fonte: Elaborado pelo autor

A partir da apresentação da imagens e códigos procurou-se facilitar a compreensão da implementação do *software*. Embora os trechos de código expostos tenham sido curtos, imagina-se que é possível entender a ideia geral da interface. Este cenário possibilita uma primeira familiarização com a interface, tornando possível discutir, a seguir, os critérios teóricos que fundamentam este estudo.

5.4 DISCUSSÃO DA INTERFACE DE RESULTADOS DE BUSCA

Pretende-se nesta seção discutir o aporte teórico das áreas de RI, IHC e Semiótica na concepção e idealização do *software*. Inicialmente serão lembradas algumas dificuldades encontradas nas interfaces de resultados de busca e alguns elementos destacados pelos estudos de comportamento informacional. Destacam-se o modelo vetorial da RI, o modelo de pensamento diagramático de Peirce e a semiótica, além dos conceitos de interação e visualização destacados na IHC.

A Seção 2.3 apresentou alguns desafios que perduram atualmente nas interfaces de RI. Singh, Hsu e Moon (2013) detectaram que em geral os resultados de busca são apresentados em listas textuais. Os autores explicam que este formato nem sempre é adequado devido à diversidade de documentos disponíveis em sistemas de busca. A sobrecarga de informações e o excesso de documentos foram abordados por alguns estudos como Case (2012), Jansen, Spink e Saracevic (2000), Dias e Carvalho (2007), Vieira e Correa (2011). Neste cenário, indicou-se que os usuários em geral focam suas atenções apenas na primeira página da lista, com cerca de 10 até 20 documentos.

Os estudos de usuários também ressaltaram fatos importantes no processo de busca. Em relatos discutidos por Pajic (2014), Pavão (2014), Huang (2007) e Lima (2003) observa-se a tendência de propor modelos alternativos de interfaces que utilizam elementos gráficos, objetivando estabelecer uma relação mais satisfatória com o usuário. Segundo Huang (2007), é interessante a criação de um ambiente comunicativo para estabelecer a conexão entre a compreensão de símbolos e a produção de significados.

Vale citar ainda as pesquisas mais atuais sobre formas alternativas de apresentação de resultados e sua relação com os usuários. Recorre-se aos estudos de Sedghi, Zeinab e Tahamtan (2018), Tomasi, Schuff e Turetken (2018), Collares et al. (2018), em que é possível destacar que embora exista certa preferência por sistemas tradicionais, os usuários atestaram interações positivas com a interface de pesquisa visual. Um dos aspectos que afeta

diretamente a utilização de uma interface é a experiência do usuário com a ferramenta. Independente da forma de abordagem, os autores estimulam o desenvolvimento de outros estudos que possam investigar formas de interação dos usuários. A implementação de forma gradativa de novos modelos de interação pode ser o caminho para que os usuários ousem utilizar inovações.

No que diz respeito à apresentação dos resultados em lista, Hobbs, Pfitzner e Powers (2002) alegam que, em geral, não é informado ao usuário os critérios pelos quais o documento foi recuperado, nem a relação do mesmo com o termo de busca. Os autores também citam a dificuldade em inspecionar longas listas e a falta de representação das semelhanças entre os documentos.

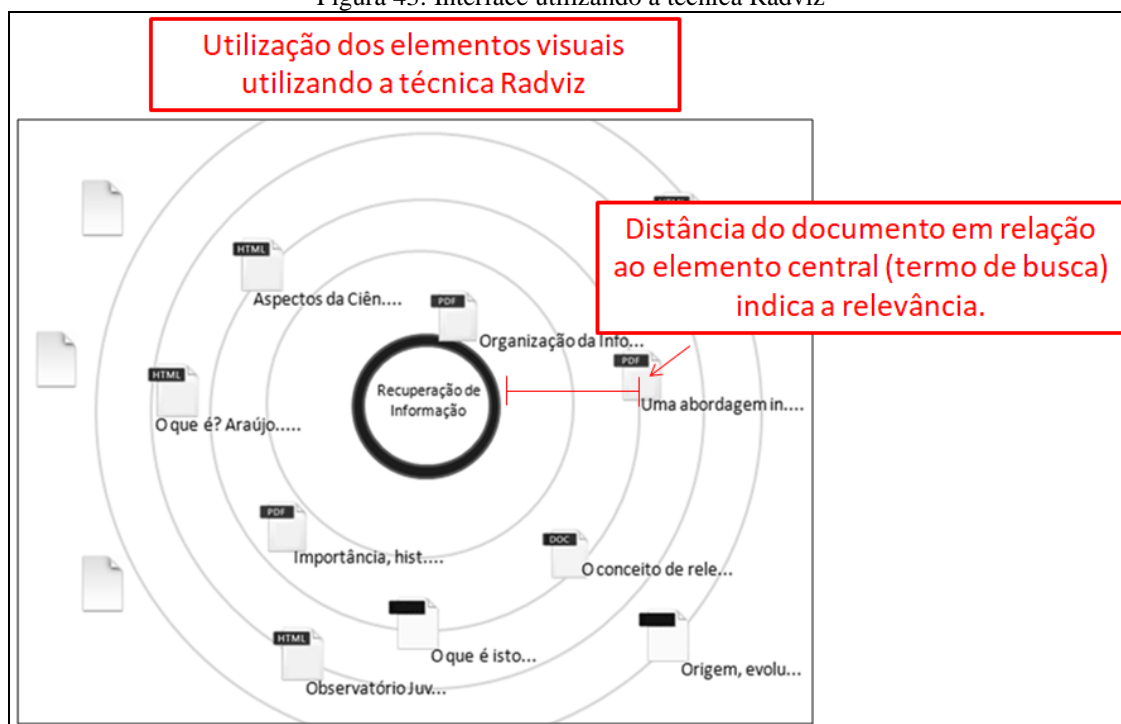
Em relação à semiótica, apoia-se na ideia do raciocínio diagramático de Peirce que foi considerado nas pesquisas de Marcondes Filho (2013), Hoffmann (2013), Farias (2012) e Inácio (2006). Esses estudos sustentam-se na visão de que a representação visual facilita a aquisição de informações em contraponto a fórmulas lógicas ou linguagens escritas. A ideia diagramática fundamenta esta perspectiva que considera os diagramas no auxílio ao processamento de informação através de especificações visuais de proximidade, forma, cor e padrões.

Através dessa conjectura, observa-se então a possibilidade da utilização de elementos visuais em formato de diagrama, conforme já apresentado na Seção 5.3. Esta proposta não visa substituir os formatos em lista amplamente utilizados e que evoluíram intensamente nos últimos anos. Pretende-se disponibilizar uma alternativa com as bases teóricas robustas da Recuperação de Informação e Semiótica Peirceana, e viável através dos conceitos da IHC e de tecnologias acessíveis utilizadas no desenvolvimento de *softwares* para internet.

5.4.1 Disposição dos elementos através da técnica Radviz e o modelo vetorial

Já apresentados os aspectos técnicos do protótipo, faz-se interessante discutir o formato de apresentação dos resultados, assim como a motivação para a utilização dos elementos. A Figura 43 ilustra o mesmo exemplo dado anteriormente, mas agora com foco na discussão das características da interface.

Figura 43: Interface utilizando a técnica Radviz



Fonte: Elaborado pelo autor

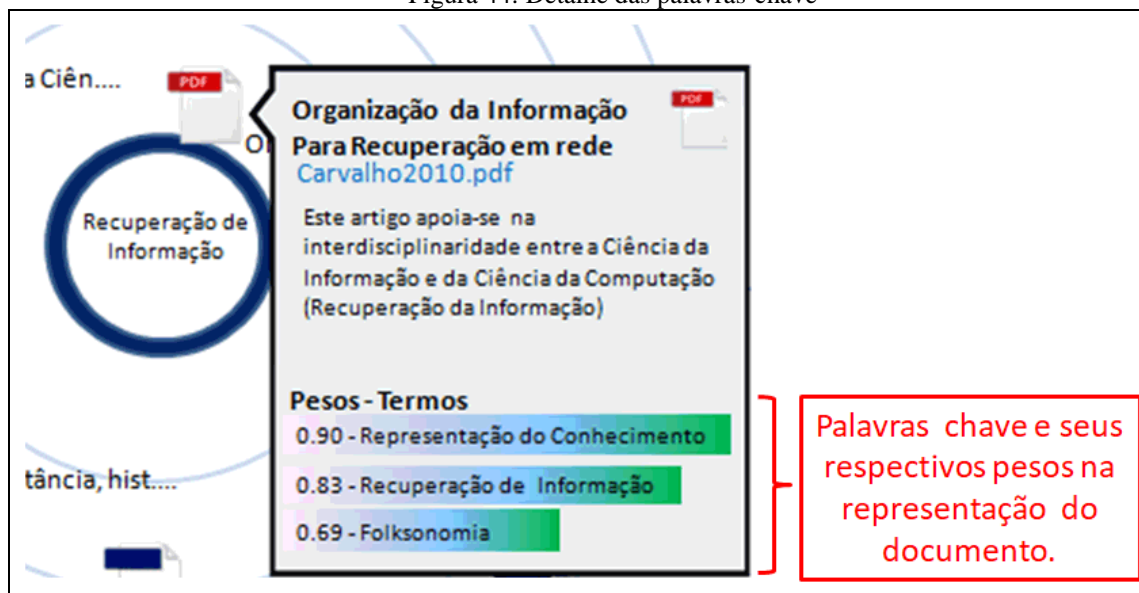
A elaboração da interface baseou-se na técnica Radviz, na qual os registros com maior similaridade ao termo de busca são visualizados próximos ao ponto central. Artero (2005) explica que essa técnica possibilita a distribuição adequada dos elementos na interface. Nesse formato são apresentadas diversas linhas que emanam radialmente do centro de um círculo e terminam no seu perímetro. O mapeamento dos pontos é uma transformação não linear que preserva algumas simetrias onde registros similares estão próximos. Para o autor, uma grande vantagem deste modelo é a baixa complexidade no processamento computacional. A Figura 43 apresenta este formato radial concêntrico em que os documentos orbitam o termo central e a proximidade representa a similaridade do documento à pesquisa realizada.

Para o posicionamento dos documentos, e por consequência o cálculo de sua similaridade, utiliza-se o modelo vetorial já embutido no Solr onde cada documento recebe uma pontuação (*score*). Quanto mais alta a pontuação, maior a similaridade do documento em relação à requisição. Conforme já apresentado, este modelo utiliza-se da ponderação TF-IDF. No caso do mapeamento para o esquema Radviz, quanto mais alto o *score* do documento, este fica posicionado mais próximo do termo de pesquisa localizado no centro da visualização.

Além da utilização do cálculo de similaridade, o modelo vetorial fornece a indexação dos documentos através de palavras-chave, e cada documento é indexado e representado pelas palavras mais relevantes ao seu conteúdo. É possível verificar nos

documentos retornados quais palavras são as mais relevantes para o próprio documento e, caso necessário, utilizá-las para novas pesquisas.

Figura 44: Detalhe das palavras-chave



Fonte: Elaborado pelo autor

As palavras-chave detalhadas na seleção do documento podem ser observadas na Figura 44. Especificamente neste exemplo, o documento é representado por três palavras que possuem pesos diferentes, variando de 0 a 1. Quanto mais próximo ao número 1, maior a importância do termo para o documento. Nota-se também que a pesquisa é feita pelo termo "Recuperação de Informação" que é o segundo termo com maior importância para o documento, entretanto ele está bem próximo ao centro da representação, o que significa que nesta coleção de documentos este documento é o que detém a pontuação mais alta. A apresentação dos principais termos que indexam o documento é importante também para possibilitar ao usuário entender o motivo pelo qual o documento foi retornado.

Em resumo, utilizam-se a técnica Radviz em conjunto com os atributos oferecidos pelo modelo vetorial (operacionalizado pelo Solr) para determinar posições que expressam a similaridade dos termos em relação aos documentos resultantes da busca. A seguir serão ressaltados através da IHC os estilos de interação presentes na interface.

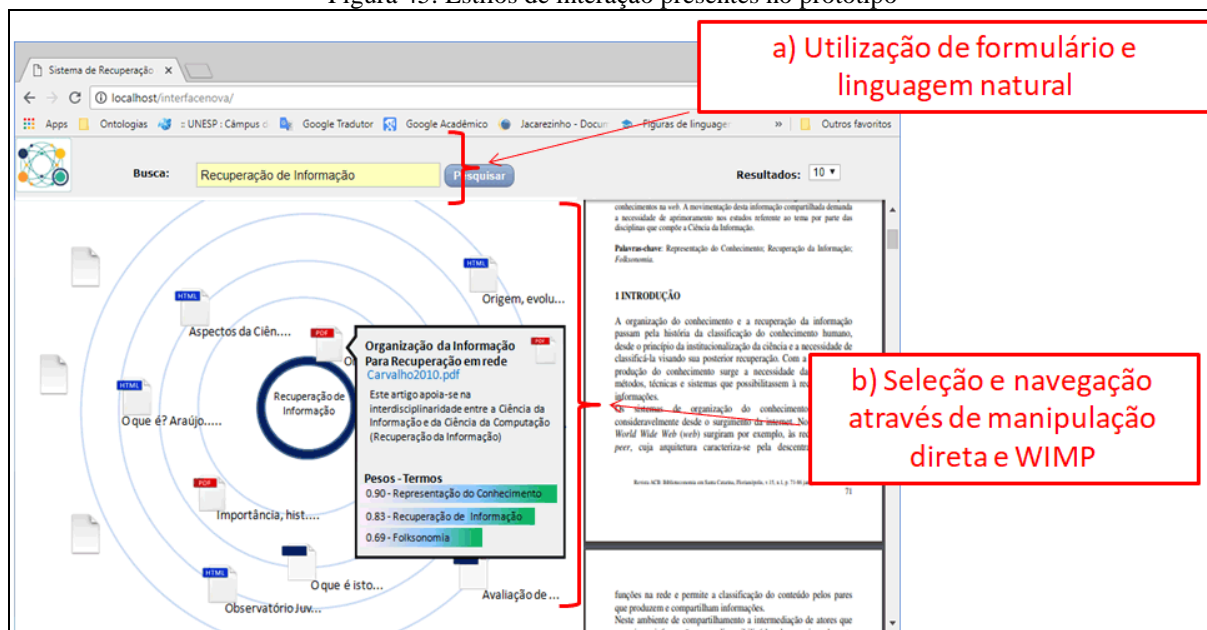
5.4.2 Estilos de interação do protótipo

Durante a apresentação da área de IHC foram elencados os seguintes estilos de interação: linguagens de comando, menus, WIMP, preenchimento de formulário,

manipulação direta e linguagem natural. Todos são utilizados em sistemas computacionais a depender das necessidades de cada *software*. Em particular nesta pesquisa são mescladas estas técnicas para proporcionar a interação dos usuários com a interface.

A Figura 45 expõe os estilos de interação disponíveis para utilização. A busca (item "a" da Figura 45) é operacionalizada através de um formulário em que o usuário utiliza a linguagem natural para expor a sua necessidade de informação. Esse formato possibilita introduzir os termos de busca no formulário e acionar a pesquisa através do botão apropriado. O item "b" da Figura 45 diz respeito à interação realizada na área de resultados. O estilo WIMP utiliza-se de apontadores visuais para indicar o movimento do mouse ou teclado por exemplo. A manipulação direta proporciona agir diretamente sobre as representações visuais para realizar as tarefas disponíveis na interface. Neste exemplo, uma possibilidade é utilizar o mouse para selecionar ou acessar o documento, agindo diretamente sobre o elemento que representa o mesmo.

Figura 45: Estilos de interação presentes no protótipo



Fonte: Elaborado pelo autor

Com a mescla dos estilos de interação, buscou-se utilizar de estruturas já conhecidas pelos usuários que acessam sistemas de busca através da pesquisa por termos. Como a proposta tem foco na utilização de elementos visuais, a manipulação direta apresenta-se como uma opção válida, visto que pode ser utilizada a partir de qualquer dispositivo computacional que acesse o sistema através de navegadores de internet.

Após discutidas as formas de interação e a estrutura para disposição dos elementos na interface, é extremamente útil verificar o aporte dos conceitos de signos a partir

da semiótica peirceana no processo de percepção e navegação no ambiente.

5.4.3 Aporte da semiótica para a concepção do protótipo

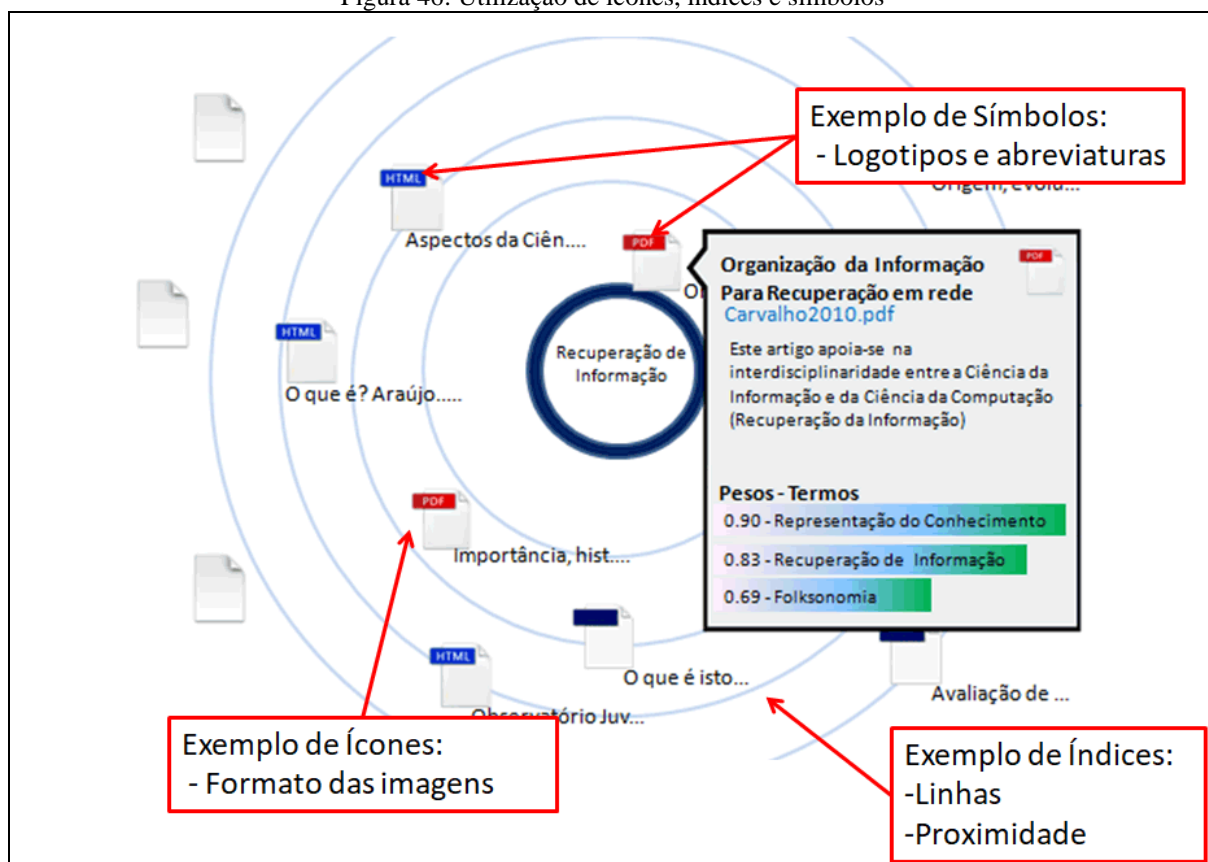
Moura (2006) discutiu as relações entre a Ciência da Informação e a Semiótica, e destacou a importância da CI compreender os aspectos sociais e técnicos envolvidos na ação de produzir, sistematizar, organizar, disseminar e recuperar informação. Refletir sobre a informação do ponto de vista semiótico proporciona um novo entendimento da realidade, provocando o aprimoramento das ações desenvolvidas pelos profissionais da informação. Em busca desse aprimoramento, verifica-se que a semiótica tem várias abordagens no estudo de interfaces digitais. Aqui serão apresentados alguns fatores detectados que influenciaram no desenvolvimento da proposta. Neste momento serão consideradas a importância do signo no processo de percepção, as inferências abduativas e a navegação em interfaces digitais. Procura-se, então, atentar-se ao potencial comunicativo verificado pelo Engenharia Semiótica, destacado por Gambarato (2013) em seu estudo sobre a percepção.

Os tipos de mediação (icônica, indicial e simbólica) se apresentam na interação humano computador através da relação dos signos com o objeto. Nesta proposta de Peirce existe a divisão em três elementos já descritos, ou seja, o ícone, o índice e o símbolo. Observa-se, então, a utilização dessas características na compreensão da utilização de alguns elementos dispostos na interface. O julgamento perceptivo apresenta as três perspectivas equivalentes a esses três tipos sógnicos: a perspectiva icônica, a indicial e a simbólica. A Figura 46 apresenta esses elementos que são utilizados na interface de resultados em uma busca por informação.

O ingrediente icônico intenciona apresentar ao usuário a percepção através da semelhança com o objeto. Na Figura 44 são usados signos no formato de uma folha de papel com uma dobra superior; o objetivo é expressar a iconicidade através da analogia deste signo com um documento textual. No que tange à dimensão indicial, o signo apresenta significado através de indicadores que o ligam à situação real, trabalhe-se com a distância ou proximidade dos objetos para indicar níveis de similaridade. As linhas radiais em volta do termo central atuam para facilitar a análise do observador sobre a distância dos diferentes documentos que se encontram posicionados no perímetro de observação. Por fim, a dimensão simbólica atua na representação por convenção, ou seja, não tem relação de semelhança ou relação indicial com o que se pretende representar, é atribuída arbitrariamente. Esta dimensão é apresentada através de signos que contêm siglas ou palavras como, por exemplo, a sigla do

formato dos arquivos.

Figura 46: Utilização de ícones, índices e símbolos



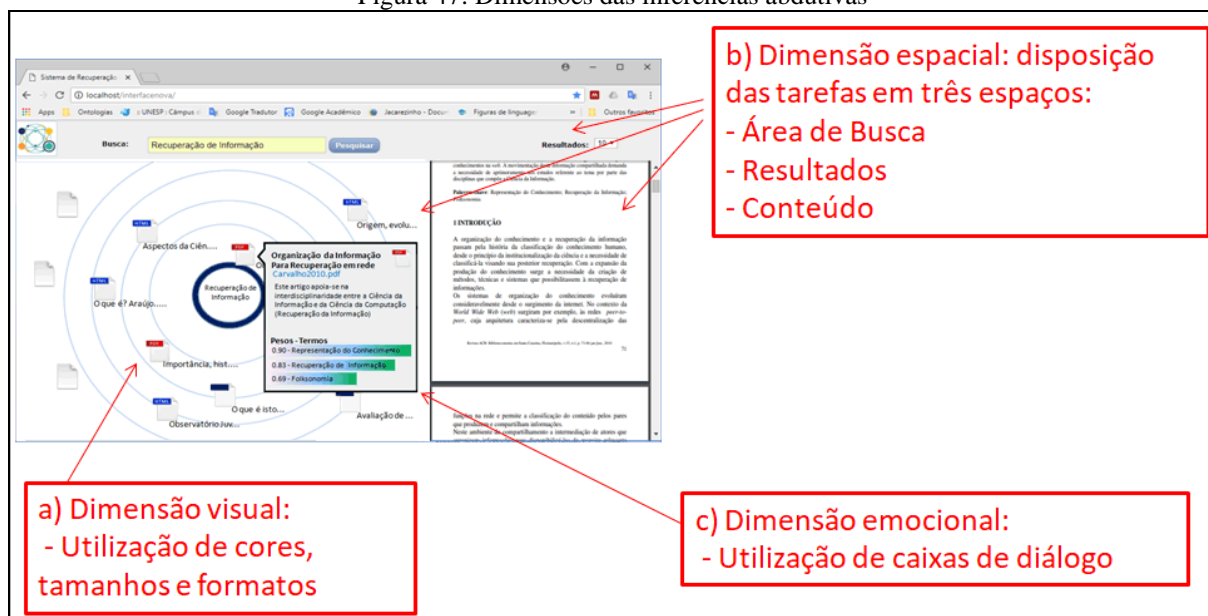
Fonte: Elaborado pelo autor

Discutindo o conceito de hibridização entre as linguagens verbais, visuais e sonoras, Monteiro (2009) argumenta que cada uma dessas linguagens tem orientação ligada às categorias designadas por Peirce. A sonora tem seu eixo baseado na iconicidade, pois quem busca necessita de conhecimento ou hábito da sintaxe sonora. A linguagem visual é uma questão indicial, devido à referencialidade. A linguagem verbal traduz a representatividade de todas as outras, devido ao seu poder de tradução. Assim, surgem os mecanismos híbridos que associam essas linguagens nos resultados de busca. A autora expõe que na combinação das linguagens "os resultados estão imbricados, hibridados e entrelaçados numa miríade de signos especialmente imagéticos e verbais, tornando a busca mais interessante e intuitiva" (MONTEIRO, 2009, p.97). É possível visualizar na Figura 46 e na Figura 47 a utilização das linguagens visuais e verbais. Considerando que toda linguagem está ligada à percepção, os meios híbridos promovem a relação de duas ou mais linguagens e/ou mídias, fornecendo estímulo ou mudanças na posição relativa dos sentidos.

No trabalho de Magnani e Bardone (2006) destacam-se as inferências

abduativas através de três dimensões: a visual, a espacial e a emocional. Revelou-se importante a atuação nessas dimensões em projetos de interfaces web. A Figura 47 expõe a dimensão visual através do uso de cores, tamanhos e formatos. A dimensão espacial é trabalhada com disposição das tarefas em áreas diferentes. Na implementação do protótipo desta pesquisa, optou-se por dividir a tela em três partes, sendo cada uma responsável por uma tarefa. A dimensão emocional considera-se através de caixas de diálogo para comunicação com o usuário. O item "c" da Figura 47 destaca uma caixa onde são apresentados ao usuário alguns detalhes do documento e permite a opção de selecionar o documento ou os termos de busca para outras consultas.

Figura 47: Dimensões das inferências abduativas



Fonte: Elaborado pelo autor

Em outra vertente da semiótica, Santaella (2004) discorre sobre os tipos de internautas de acordo com suas características de navegação: o errante, o detetive e o previdente. Como já especificado, a interação por linguagem natural é uma opção para internautas os três tipos de internautas, cada qual com suas especificidades. Na interface desenvolvida é possível, após uma primeira consulta, verificar nos resultados outros caminhos proporcionados pelas palavras-chave contidas nos documentos. A autora explica que o internauta detetive navega utilizando indícios (pistas) que são disponibilizados nesta proposta, conforme já descrito na Figura 46. O internauta previdente já é mais experiente e pode prever o resultado de suas ações e, nesta proposta, também utiliza-se dos sinais para realizar suas buscas.

Através dos pontos ressaltados, justifica-se a utilização da semiótica como

base teórica no desenvolvimento desta proposta. Considerando a semiótica peirceana, pode-se entender como os diferentes processos de percepção humana podem influenciar na utilização de elementos que visam proporcionar alternativas para interfaces *web*.

5.5 USABILIDADE

A avaliação de interface é uma etapa importante no processo de desenvolvimento de aplicações computacionais. Hartson (1998) argumenta que ao avaliar é possível estimar o sucesso ou o insucesso da interface em relação aos requisitos e objetivos propostos. Mesmo que o desenvolvimento se apoie em diretrizes e princípios robustos, é inevitável avaliar o resultado. A ideia de realizar avaliações iniciou-se juntamente com o princípio dos estudos de IHC. Nielsen (1993) foi um dos precursores do estudo sobre usabilidade. Para o autor, a realização de testes com usuários reais é um método de usabilidade fundamental e, em certo sentido, insubstituível, pois fornece informações diretas sobre como as pessoas usam computadores e quais são seus reais problemas com a interface concreta sendo testada.

Rocha e Baranauskas (2003) apresentam três grandes objetivos da etapa de avaliação: avaliar a funcionalidade do sistema, avaliar o efeito da interface junto ao usuário e identificar problemas específicos do sistema. Percebendo a necessidade e os objetivos da avaliação, é importante selecionar os procedimentos que farão parte do processo. Santos (2016) e Nascimento e Amaral (2010) destacam e classificam as principais técnicas de avaliação:

1. Técnicas prospectivas: baseiam-se na opinião dos usuários, em geral realizadas por meio de entrevistas e questionários.
2. Técnicas preditivas: baseadas em conhecimento teórico do avaliador, que é um especialista. Podem ser realizadas através de uma avaliação heurística ou utilizando listas de verificação.
 - a) Avaliação Heurística: realizada por especialistas e baseada em regras de âmbito geral e critérios ergonômicos, como por exemplo: adaptabilidade, compatibilidade, confiabilidade, entre outros;
 - b) Lista de verificação (*checklist*): são vistorias a partir de listas com itens a serem verificados; não são necessariamente realizadas por especialistas em que cada item é selecionado

caso a interface atenda à especificação.

3. Técnicas objetivas: são também conhecidas como empíricas e baseadas na participação direta dos usuários.

a) Ensaio de Interação: consiste em "ensaiar" a utilização de um sistema, ou seja, são realizadas sessões onde as pessoas realizam as tarefas especificadas.

De acordo com Preece, Rogers e Sharp (2005), as técnicas podem ser categorizadas e aplicadas de várias formas, podendo relacionar mais de uma, a depender dos objetivos que podem ser: solicitar opinião de especialistas, observar, coletar opiniões e testar o desempenho dos usuários. As definições de metas e questões são necessárias no processo de avaliação, portanto ao definir a técnica de avaliação é importante que esses aspectos estejam contemplados no início do processo. Para orientar o processo de avaliação, propõe-se utilizar o esquema DECIDE proposto por Preece, Rogers e Sharp (2005), formado pelas iniciais das palavras: *determine, explore, choose, identify, decide, evaluate* (determine, explore, escolha, identifique, decida, avalie). Essas etapas são descritas da seguinte forma:

- a) **Determinar** os objetivos gerais que a avaliação deverá tratar.
- b) **Explorar** perguntas específicas a serem respondidas. Delimitar os objetivos gerais em perguntas específicas ao sistema a ser avaliado.
- c) **Escolher** o paradigma e as técnicas de avaliação que responderão às perguntas elencadas na etapa anterior.
- d) **Identificar** questões práticas que devem ser tratadas como: perfil e número de usuários; ambiente em que a avaliação será realizada; tarefas; planejamento e preparação do material de avaliação e recursos e equipamentos para a realização da avaliação.
- e) **Decidir** como lidar com questões éticas. Certificar-se de que o direito das pessoas que participam do teste está sendo respeitado.
- f) **Avaliar**, interpretar e apresentar os dados, apresentando a discussão dos resultados.

Com base nos estágios descritos, é possível estabelecer uma estrutura adequada para o processo de avaliação. O desenvolvimento das próximas seções seguirá esta estrutura. Embora as etapas estejam intimamente interligas e dependentes, elas serão agrupadas com o objetivo de favorecer o desenvolvimento do texto. Inicialmente serão discutidos os três primeiros itens (determinar, explorar e escolher), delimitando o que será avaliado e quais ferramentas serão utilizadas para esta aplicação. Em sequência serão

contempladas a quarta e quinta etapas (identificar, decidir) em que, além de definições da estrutura do teste, serão verificados os aspectos éticos. Por fim, será aplicado o teste (etapa avaliar) e elaborada a discussão dos seus resultados.

5.5.1 Delimitação do teste de Usabilidade: determinar, explorar e escolher

Esse tópico tem como intuito indicar os passos iniciais do plano de teste e aqui serão definidos o objetivo geral, questões a serem respondidas pela interface e as ferramentas a serem utilizadas durante a execução dos testes. Durante a definição dos objetivos, será realizada uma conexão com os requisitos da aplicação descritos pelo diagrama de casos de uso, ilustrado no item 5.1.1. Como a avaliação gira em torno da interação do usuário com a interface, os casos de uso que determinam essas ações são: realizar busca, selecionar documento e acessar documento. Sendo assim, fica estabelecido que o objetivo geral deste plano de teste é avaliar a interação do usuário em relação aos requisitos da aplicação.

As funcionalidades representadas pelos casos de uso foram geradas a partir da identificação de dificuldades encontradas em interfaces de recuperação de informação. Dessa forma, cada caso de uso deve proporcionar percepções que atendam a tais indicações. A partir desse paralelo é possível designar algumas questões específicas a serem correspondidas pela interface.

Quadro 8: Percepções esperadas para cada caso de uso

Caso de Uso	Indicação	Questões Específicas
Realizar busca	Proporcionar a identificação dos documentos mais relevantes a partir da interface de resultados	É possível identificar os documentos mais relevantes a partir da interface proposta?
Selecionar documento	Fornecer uma visão detalhada do documento indicando as palavras-chave mais representativas do seu conteúdo.	O usuário tem acesso a detalhes do documento sem necessitar abrir o documento completo? A indicação das palavras-chave mais pertinentes ao documento está disposta de forma clara?
Acessar documento	Possibilitar o acesso ao documento na íntegra.	A opção de acessar o documento integral está adequada?

Fonte: Elaborado pelo autor

O Quadro 8 relaciona as principais questões abordadas no plano de teste e que servirão como base para a análise e discussão dos resultados. Torna-se então necessário definir as técnicas e ferramentas que auxiliarão o processo de coleta de informações dos usuários. Conforme indicado nas pesquisas de Santos (2016), Nascimento e Amaral (2010),

Preece, Rogers e Sharp (2005), Rocha e Baranauskas (2003), Hartson (1998) e Nielsen (1993), a realização da avaliação com usuários é um método fundamental de usabilidade, pois proporciona analisar fatores de interação em situações reais de uso.

Com base nessa premissa e na consideração da necessidade de identificar os fatores perceptivos dos usuários, decidiu-se utilizar o ensaio de interação. Essa escolha foi feita por ser esse um método flexível em relação ao perfil dos experimentadores envolvidos e, principalmente, por envolver usuários. Nascimento e Amaral (2010) salientam que este tipo de teste permite a coleta de dados qualitativos e/ou quantitativos a partir da coleta de dados por meio de observações de interação entre usuários e sistema.

Para operacionalizar a coleta de informações é indispensável o uso de ferramentas que forneçam dados sobre as interações. Santos (2016) indica a utilização de cenários que consistem em conjuntos de tarefas que devem ser realizados pelos usuários. Essas tarefas são selecionadas a partir de algumas premissas: objetivos do *software*, requisitos, além das funcionalidades mais acionadas e fundamentais do *software*. O autor comenta que com a definição das tarefas, os testes são realizados no ambiente do usuário ou em laboratórios. Após a aplicação do teste, a satisfação do usuário é avaliada utilizando-se questionários e entrevistas. Além disso, Preece, Rogers e Sharp (2005) apresentam outras formas de coleta de dados, como *logs* de interação e gravação em vídeo, por exemplo. Com base nessas indicações, esta pesquisa seguirá o seguinte percurso: definição do cenário e das tarefas, definição do questionário, aplicação do teste, aplicação de questionário e análise de *log* de interação.

O cenário elaborado para o teste da proposta desta pesquisa baseia-se nos objetivos da interface e nas questões levantadas e exibidas no Quadro 8. Desta forma, criou-se um cenário com cinco tarefas como disposto no Quadro 9 e discriminado no apêndice A.

Esta proposta objetiva proporcionar ao usuário a experiência de uso do *software*. Após a execução de cada tarefa, responde-se à seguinte questão "Foi possível realizar a tarefa?". Como resposta, são indicadas uma das opções a seguir: "sim, com facilidade"; "sim, com dificuldade"; "não conseguiu realizar a tarefa". Além disso, são anotadas observações sobre possíveis dificuldades.

Quadro 9: Cenário para o roteiro de teste

CENÁRIO: REALIZAR BUSCA BASEADA EM PALAVRAS-CHAVE		
Tarefa	Interface relacionada	Objetivo/Questões
1. Você precisa realizar uma pesquisa baseada em palavras-chave, assim como faz no Google. Seu objetivo é explorar o termo "Recuperação de Informação". Informe o termo de pesquisa na interface.	Figura 35	Realizar a busca por palavras-chave
2. Analisar os resultados da busca e indicar qual o documento se apresenta como mais relevante. Qual arquivo corresponde à busca realizada?	Figura 37	É possível identificar os documentos mais relevantes a partir da interface proposta?
3. Como você faria para saber detalhes sobre o documento de seu interesse?	Figura 40	O usuário tem acesso a detalhes do documento sem necessitar abrir o documento completo?
4. Como você faria para acessar o documento mais relevante?	Figura 41	A opção de acessar o documento integral está adequada?
5. No detalhamento do arquivo aparecem alguns termos na parte inferior, qual sua percepção sobre esses itens?	Figura 40	A indicação das palavras-chave mais pertinentes ao documento está disposta de forma clara?

Fonte: Elaborado pelo autor

Após a realização das tarefas dispostas no cenário, é entregue ao usuário um questionário para avaliar o grau de satisfação. Preece, Rogers e Sharp (2005) explicam que os questionários constituem técnicas bem estabelecidas para coletar informações sobre os usuários e dados demográficos. Algumas informações demográficas básicas (exemplo: gênero, idade) e de experiência (exemplo: tempo de uso de computadores, experiência em busca por informação) são úteis para conhecer a diversidade no grupo de amostragem. Após essas questões mais genéricas, são realizadas perguntas mais específicas para apoiar as metas da avaliação.

Para elaboração das questões foram consideradas algumas recomendações elencadas por Preece, Rogers e Sharp (2005), dentre as quais destacam-se: realizar perguntas claras e específicas, evitar questões complexas, oferecer perguntas fechadas com várias possibilidades, utilizar questionários compactos e indicar instruções claras nas questões sobre escalas ou marcação de um "x" por exemplo. Os autores também apresentam tipos diferentes de respostas, a depender das perguntas. O formato das respostas varia de um "sim" ou um "não", até o limite de proporcionar ao usuário faixas de possibilidades. Nesta pesquisa serão utilizadas questões de escolha única para coletar dados demográficos e de experiência do usuário.

Não avaliação de opinião dos usuários sobre questões relacionadas ao

software, serão utilizados dois modelos de escalas: escala Likert e escalas de diferencial semântico. Para Preece, Rogers e Sharp (2005) são formatos válidos e adequados quando se necessita conhecer o julgamento sobre aspectos da interface. A escala Likert é utilizada para medir opiniões, crenças e atitudes, por isso são frequentemente utilizadas para avaliação de satisfação dos usuários. Em geral é utilizada uma faixa de valores (de 1 a 5 por exemplo), ou palavras como: concordo totalmente, concordo, sem opinião, discordo e discordo totalmente. As escalas de diferencial semântico exploram faixa de atitudes bipolares a respeito de um item, onde cada par de atitudes é representada por um par de adjetivos. O usuário deve indicar um "x" em uma das posições entre os dois extremos, a fim de indicar sua concordância com tal adjetivo, como demonstra a Figura 48.

Figura 48: Exemplo de escala de diferencial semântico

		3	2	1	0	-1	-2	-3		
11. Avalie a sua percepção em relação ao uso da interface. Assinale com um "x" quando acredita que o atributo representa sua opinião. Deve ser assinalado apenas um "x" em cada linha										
Simple										Complexo
Útil										Inútil
Intuitiva										Não intuitiva

Fonte: Elaborado pelo autor

O questionário apresenta-se como técnica bastante útil no processo de avaliação de interação. Através dele é possível obter informações importantes sobre o perfil do usuário, suas dificuldades de interação com o sistema e suas sugestões. A elaboração do questionário utilizado neste plano de testes contempla a apreciação sobre a opinião dos usuários em relação às tarefas anteriormente relacionadas. Para isso, em determinados momentos foram descritas duas perguntas com o mesmo objetivo de análise, mas com enfoques diferentes, conforme exemplificado no Quadro 10. Preece, Rogers e Sharp (2005) argumentam que em determinadas situações, a mudança de direção das perguntas ajuda a verificar as intenções dos usuários. O questionário completo está disponível no apêndice B.

Quadro 10: Exemplo de perguntas do questionário de avaliação

Meta a responder	Perguntas elaboradas	Opções de resposta
É possível identificar os documentos mais relevantes a partir da interface proposta?	Conseguir identificar o documento mais relevante indicado pela interface.	Escala de 1 a 5, onde: 1 - discordo totalmente 5 - concordo totalmente
	Não foi possível identificar qual documento era o mais relevante na busca.	1 - discordo totalmente 5 - concordo totalmente

Fonte: Elaborado pelo autor

Além dos questionários e da definição das tarefas, outra ferramenta utilizada é a análise de *log*. Essa técnica consiste em registrar interações com mouse, teclado e outros dispositivos para identificar as interações dos usuários. Especificamente neste plano de teste será utilizada uma biblioteca feita em JavaScript denominada "clickheat" que mapeia os cliques do *mouse* que o usuário realiza na interface. Com base nessas informações é possível verificar as áreas em que o usuário utiliza tal função e interage com elas. A partir dessa coleta é gerado um mapa de calor que será apresentado na seção de testes.

Nesta seção foram delimitados os objetivos deste plano de teste, com isso foi possível selecionar as ferramentas e elaborar os cenários e questionários a serem utilizados no teste de usabilidade. A seguir será estruturado o plano de teste.

5.5.2 Estruturação do teste: "identificar" e "decidir"

A etapa "identificar" consiste em abordar questões práticas em relação à aplicação dos testes em relação aos usuários, ambiente de teste e recursos. Por sua vez, a etapa "decidir" visa assegurar que os direitos das pessoas que participam do teste estão sendo assegurados.

No que diz respeito aos usuários, foram recrutados de forma voluntária e aleatória. A não definição de um perfil específico faz parte da ideia do *software* de atender a públicos heterogêneos. Para resguardar os direitos dos participantes, além do esclarecimento dos objetivos e formato dos testes, os questionários foram respondidos de forma anônima, consequentemente não é possível associar os dados coletados aos usuários. Para isso foram seguidas as seguintes recomendações:

- a) Informar aos participantes os objetivos, indicando que a avaliação é do *software* e não do usuário. Esclarecimentos sobre tempo previsto e quais dados serão coletados.
- b) Alertar que os usuários estão livres para terminar o teste a qualquer

momento.

- c) Indicar que os dados coletados não contêm informações que possibilitem a identificação do usuário.

Para certificar essas recomendações, foi elaborada uma carta de consentimento (apêndice c) que traz esclarecimento sobre essas situações. Assim verifica-se o cuidado em preservar a identidade dos autores e garantir o seu voluntariado no processo. O recrutamento foi realizado através de divulgação verbal em uma instituição de ensino médio técnico e aberto aos estudantes e servidores, dos quais foram selecionados os quatro primeiros voluntários de cada grupo, tendo em vista ainda possíveis desistências.

Optou-se por uma quantidade reduzida de usuários fundamentando-se nas indicações de Nielsen (1993) e Nielsen (2000). O autor baseou-se em pesquisas anteriores e fatores estatísticos para indicar que a realização de testes com cinco usuários seria suficiente para realizar diagnósticos na interface. Isso se explica, uma vez que as novas descobertas vão decrescendo, gradativamente, à medida que se observa os usuários, ou seja, o primeiro usuário observado proporciona a coleta de apontamentos que se repetem nos subsequentes, por consequência, a cada teste executado, a probabilidade de surgirem novas descobertas diminui. Neste ponto, o autor descreve que a quantidade de cinco indivíduos já proporciona um nível satisfatório de precisão. A quantidade de participantes da pesquisa é uma amostra restrita, entretanto atendeu aos objetivos iniciais dessa fase do desenvolvimento da interface que buscou realizar a primeira fase de testes. Obviamente que para a aplicação de forma comercial da ferramenta devem ser realizados outros testes com uma amostra mais robusta.

A partir da definição dos participantes, apresenta-se no Quadro 11 os recursos necessários e utilizados nos testes.

Quadro 11: Recursos para o plano de testes

Definição	Operacionalização
Local e período de aplicação do teste	Laboratório de informática no período de dezembro de 2018 a janeiro de 2019
Previsão do tempo de teste	Aproximadamente 30 minutos
Suporte computacional necessário	Computador pessoal com sistema operacional Windows 10 e navegador Google Chrome
Outros <i>softwares</i> necessários	Servidor Apache PHP Solr
Número de usuários testados	8

Fonte: Elaborado pelo autor

O ambiente foi preparado a partir dos requisitos discriminados. O

laboratório de informática foi escolhido por ser um local ao qual os usuários já estão ambientados. Devido à proposta desta pesquisa estar em desenvolvimento, portanto ainda sem sua implementação completa, o teste mescla o protótipo básico e uma simulação interativa. O protótipo está vinculado às ações já implementadas na interface, como a busca e a exibição dos resultados em formato diagramático. Para complementar e exibir aos usuários a função do detalhamento do conteúdo dos documentos, são simuladas imagens previamente elaboradas que proporcionam apresentar o formato final da interface.

5.5.3 Aplicação e análise: avaliar

Serão apresentados os experimentos realizados para analisar a abordagem proposta. Esta é a última fase e se realiza a partir dos testes de usabilidade já delineados no processo descrito pelo framework DECIDE. Com base nos objetivos e questionamentos levantados, é utilizado o ensaio de interação diretamente no protótipo da interface. Para coleta de dados, os usuários são observados, respondem a um questionário e as interações com o *mouse* são monitoradas a partir da biblioteca *clickheat* que foi associada à interface.

5.5.3.1 Perfil dos usuários e execução das tarefas

Como descrito anteriormente, os experimentos foram realizados por oito participantes recrutados de forma voluntária entre estudantes e servidores de uma instituição de ensino médio/técnico. Uma síntese do perfil dos participantes está disposta no Quadro 12.

Quadro 12: Síntese do perfil dos participantes

Característica	Informações
Idade	- 4 participantes de 14 a 23 anos - 1 participante de 24 a 33 anos - 2 participantes de 34 a 43 anos - 1 participante de 44 a 53 anos
Média de idade	23 anos
Sexo Masculino	3 participantes
Sexo Feminino	5 participantes
Conhecimentos em Informática	3 participantes conhecimento ótimo 3 participantes conhecimento bom 2 participantes conhecimento regular
Com que frequência utiliza os serviços de busca na Web	7 participantes usam diariamente 1 participante usam semanalmente
Experiência em ferramentas de busca	4 participantes - 8 anos ou mais 2 participantes - 3 a 5 anos 2 participantes - menos de 2 anos
Principal uso dos serviços de busca	4 participantes - atividades profissionais 3 participantes - lazer 1 participante - fins acadêmicos

Fonte: Elaborado pelo autor

Verifica-se uma faixa etária variada, com predominância de participantes com até 23 anos, todos estudantes. A maioria dos participantes indicou seus conhecimentos em informática de nível bom e ótimo. Ao todo, sete participantes indicaram que utilizam mecanismos de busca diariamente, portanto estão familiarizados com a utilização desse serviço. Sobre o motivo de utilização, três estudantes indicaram como principal uso o lazer e um indicou para fins acadêmicos. Já entre os servidores, todos indicaram a utilização para fins profissionais. Importante destacar que foi solicitado que assinalassem apenas o principal motivo, embora todos os usuários tenham mencionado o uso menos frequente para as outras situações.

Quanto à experiência de uso, 50% indicaram que utilizam há mais de 8 anos (todos na faixa etária acima de 24 anos), os outros se dividiram entre: 25% de 3 a 5 anos e 25% menos de 2 anos (os estudantes). Quanto ao mecanismo mais utilizado, 100% dos participantes apontaram utilizar o Google em suas buscas. Esse é um número expressivo, o que indica a influência desta ferramenta no que diz respeito à busca por informação.

Durante a observação do teste, verificou-se que o tempo gasto para realização das tarefas variou entre 2 e 4 minutos, obtendo-se uma média de 3 minutos. A resposta ao questionário variou de 2 a 4 minutos, alcançando uma média de 3,25. O tempo total de duração do experimento variou entre 15 e 20 minutos, contanto o contato inicial, com explicitação dos objetivos e conscientização sobre o formato das atividades e uso dos

resultados. Em geral, o tempo gasto pelos participantes foi curto e os fatos que contribuíram para esta situação são a familiaridade com mecanismos de busca, a quantidade de tarefas reduzida e o formato compacto e objetivo do questionário.

A execução das tarefas proporcionou criar uma tabela de efetividade que consiste em apresentar de forma visual a quantidade de tarefas concluídas com sucesso, concluídas com dificuldade e as que não foram concluídas. Na Figura 49 está exposta tal situação. Atendendo ao suporte apresentado pela semiótica, relaciona-se a interação dos participantes ao conceito de tipos de internautas apresentados por Santaella (2004). Observou-se nos experimentos as características apresentadas pela autora, desde a facilidade de utilização de linguagem natural pelos internautas errantes, a navegação através de indícios do internauta detetive, e a experiência do internauta previdente. Essas características ficaram evidentes, principalmente pela análise do tempo gasto nas tarefas, pois constatou-se que os usuários experientes concluíram as tarefas em tempo inferior. A exceção foi o primeiro participante que concluiu as tarefas com extrema facilidade e rapidez, embora não esteja na faixa de usuários mais experientes, mas tem um tempo adequado, pois se encaixa na faixa de 3 a 5 anos de uso de serviços de busca.

Figura 49: Tabela de execução das tarefas

	Usuário 1	Usuário 2	Usuário 3	Usuário 4	Usuário 5	Usuário 6	Usuário 7	Usuário 8
Tarefa 1: realizar uma pesquisa baseada em palavra-chave	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso
Tarefa 2: indicar o documento mais relevante	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso
Tarefa 3: analisar detalhes do documento	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso
Tarefa 4: acessar o documento na íntegra	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso	Concluiu a tarefa com sucesso
Tarefa 5: analisar palavras-chave que identificam o documento	Concluiu a tarefa com sucesso	Concluiu a tarefa com dificuldade	Não concluiu a tarefa	Não concluiu a tarefa	Concluiu a tarefa com sucesso	Concluiu a tarefa com dificuldade	Concluiu a tarefa com dificuldade	Concluiu a tarefa com dificuldade

Concluiu a tarefa com sucesso

Concluiu a tarefa com dificuldade

Não concluiu a tarefa

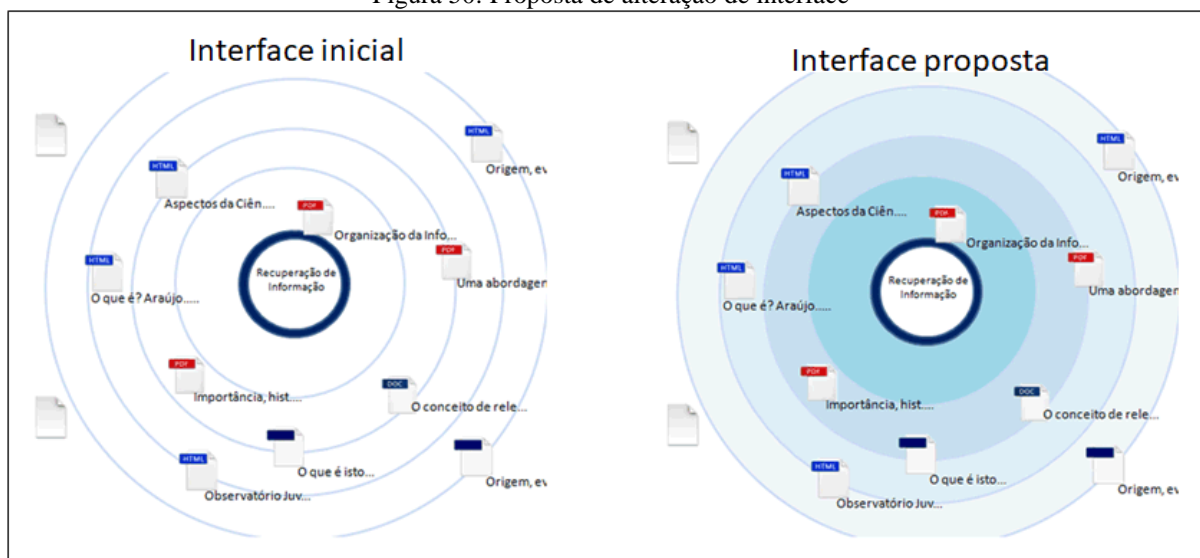
Fonte: Elaborado pelo autor

O cenário apresentou uma visão geral das tarefas desejadas para o ambiente, como pode ser observado na Figura 49. As quatro primeiras tarefas foram concluídas totalmente, com relativa desenvoltura dos participantes. A pesquisa por palavras-chave foi realizada de forma extremamente natural. Infere-se, assim, que esse tipo de interação é utilizado amplamente pelos participantes de forma frequente.

A Tarefa 2 consistia em indicar qual o documento mais relevante na visualização cumprida com sucesso. Os participantes indicaram que a proximidade do elemento com o termo de busca foi fundamental para essa percepção. Interessante associar essa ideia ao conceito de julgamento perceptivo, principalmente no que diz respeito à dimensão indicial, onde verifica-se que o signo representa indicadores que, neste caso, utiliza-se a proximidade dos objetos para indicar níveis de similaridade. O primeiro participante fez uma contribuição bastante significativa no que diz respeito à visualização de proximidade, quando foi sugerido a utilização de cores em escala para delimitar as órbitas e, conseqüentemente, melhorar a percepção de proximidade dos documentos em relação ao termo central. A

Figura 50 apresenta a projeção de tal sugestão.

Figura 50: Proposta de alteração de interface



Fonte: Elaborado pelo autor

Embora inicialmente não estivesse previsto no teste, a sugestão foi apresentada aos outros participantes. Todos indicaram aceitação da proposta e, por consequência, recomendaram a sua adoção. Conecta-se isso ao trabalho de Magnani e Bardone (2006) em que foi destacado o papel importante das inferências abduativas, em particular a dimensão visual que, entre outras situações, sugere o uso de cores para favorecer a percepção dos usuários.

As Tarefas 3 e 4 consistiam em verificar o detalhamento dos documentos e, por consequência, acessá-los. Foram executadas com fluência pelos participantes e presume-se que essa situação se deve à familiaridade que os participantes demonstraram no uso de sistemas computacionais. Remete-se à fundamentação proporcionada pela IHC que destaca os

estilos de interação, especificamente o uso do mouse, amplamente disseminado pelo estilo WIMP (apontadores visuais para indicar o movimento do mouse) e da manipulação direta que proporciona agir diretamente sobre as representações visuais para realizar as tarefas.

Por fim, a Tarefa 5 não teve o resultado esperado, uma vez que apenas dois participantes perceberam a função das palavras que identificaram os documentos e que existia a designação do grau de importância das palavras-chave ao conteúdo dos documentos. A maioria dos participantes percebeu que as palavras tinham alguma correspondência com o conteúdo do documento, mas não souberam descrever qual tipo de relação. Vê-se como necessária a utilização de outros meios para associar este conceito ao conteúdo documental.

Um possível caminho a ser explorado nessa perspectiva é a convergência entre a semiótica e as linguagens documentárias. Lara (1993) abordou a semiótica a partir da discussão sobre as linguagens documentárias, sendo o principal objetivo desse trabalho analisar os problemas de representação de informação em documentação. Já em Lara (2006), tratou-se a informação (no contexto da Ciência da Informação) como um signo elaborado com a intenção de atuar como elemento de comunicação documentária. Ao considerar os documentos como unidades textuais organizadas sob uma forma e um conteúdo, indica-se a constituição de uma unidade de interpretação. Moura, Silva e Amorim (2002) investigaram as potencialidades dos estudos da semiótica e da semiologia para a construção e utilização de linguagens de indexação. Essas abordagens se configuram como possibilidades que devem ser analisadas de maneira mais aprofundada e indicam um percurso a ser explorado.

5.5.3.2 Análise do questionário de avaliação de usabilidade e *logs* de interação

O questionário para avaliação de usabilidade é um instrumento que irá subsidiar as discussões dos testes sobre as impressões do usuário em relação à interface proposta. Após a realização do experimento, os usuários respondem a uma série de afirmações apontando sua opinião, utilizando escala de concordância e discordância. Foram adicionadas outras afirmativas que tratam de questões gerais do uso do sistema. A partir das opiniões assinaladas, foi gerada uma escala de concordância para auxiliar a apreciação dos dados, tendo sido aferida a média da opinião dos usuários que utilizaram o número 1 para discordância total e 5 para concordância. Para agrupar as afirmativas, utiliza-se as tarefas delineadas e executadas pelos usuários. Essa relação é sintetizada no Quadro 13.

Quadro 13: Relação de afirmativas em relação às tarefas.

Tarefas	Afirmativas	Nível de concordância
Tarefa 1: realizar uma pesquisa baseada em palavras-chave	1. Consegui realizar as buscas e acessar os documentos na interface disponibilizada.	4,63
Tarefa 2: indicar o documento mais relevante	2. Consegui identificar o documento mais relevante indicado pela interface.	4,88
	11. Não foi possível identificar qual documento era o mais relevante na busca.	1,25
Tarefa 3: analisar detalhes do documento	3. A disposição dos elementos permite acessá-los rapidamente.	4,88
	7. Os elementos dispostos na interface proporcionaram acesso às informações dos documentos.	4,88
Tarefa 4: acessar o documento na íntegra	5. Não consegui acessar o documento que selecionei.	1,50
Tarefa 5: analisar palavras-chave que identificam o documento	10. Consegui visualizar as palavras que representam o documento e seu grau de importância.	3,63
Questões Gerais	4. Os recursos de navegação (ícones e botões) estão todos claros e fáceis de achar.	4,50
	6. Eu achei a ferramenta complexa. Precisei pensar muito para completar as tarefas.	1,13
	8. Eu precisaria de apoio de uma pessoa para usar este aplicativo.	1,88
	9. Eu me senti muito confiante usando esta ferramenta.	3,38
	12. Essa ferramenta seria útil nas minhas atividades de pesquisa.	4,38

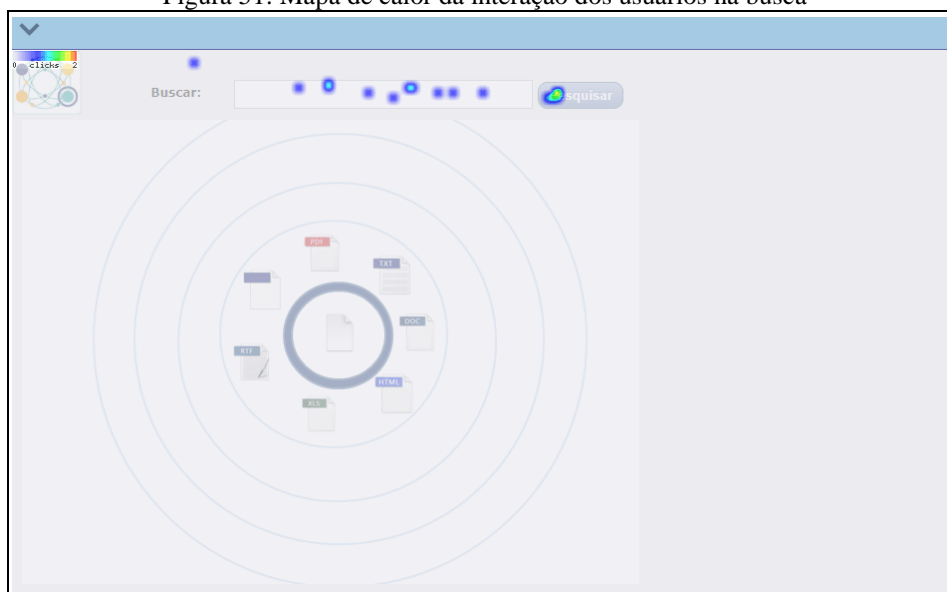
Fonte: Elaborado pelo autor

A Tarefa 1 consistiu em verificar a primeira ação dos usuários em relação ao sistema proporcionar a realização de uma busca no sistema. As respostas tiveram um alto grau de concordância quanto ao sucesso da realização da tarefa. A familiaridade dos usuários com o formato que utiliza a caixa de texto, além da simplicidade de expressar sua busca com palavras-chave, contribuíram para tal resultado. Para ratificar essa indicação, a Figura 51 apresenta o mapa de calor que considera os cliques do mouse nas primeiras interações com a interface.

Como pode ser observado, as interações estão dispostas na área que comporta o retângulo de busca e o botão que aciona a pesquisa. Houve uma situação verificada durante a realização dos experimentos, pois nem todos os participantes utilizaram o mouse para realizar a busca; alguns pressionaram a tecla "enter" após especificar o termo desejado. Mesmo tendo essa outra opção, o mapa ilustra a preferência pelo local adequado de

interação.

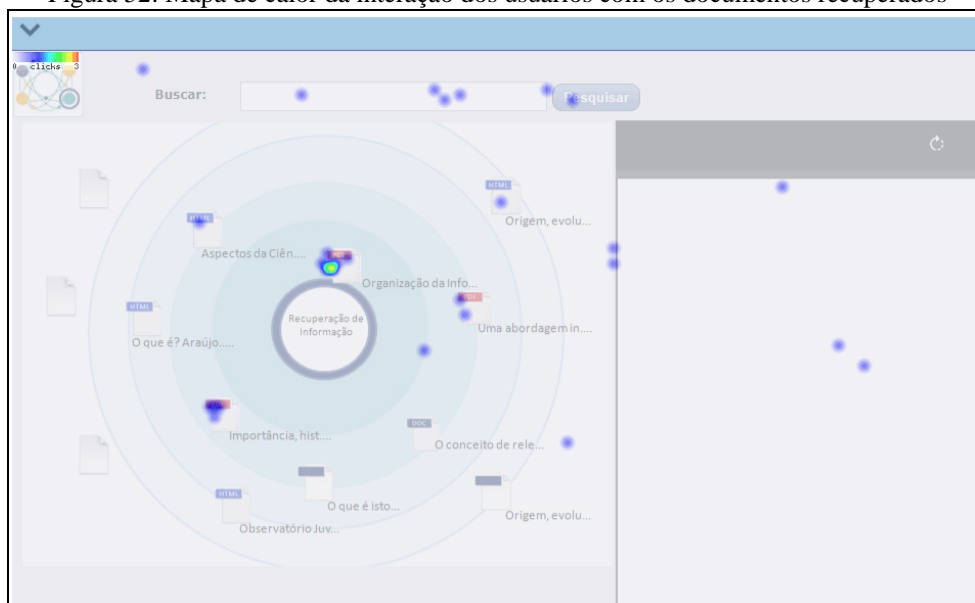
Figura 51: Mapa de calor da interação dos usuários na busca



Fonte: Elaborado pelo autor

A Tarefa 2 compreende a verificação de percepção dos documentos mais relevantes. Utilizou-se neste caso de duas afirmações. Na primeira obteve-se um nível alto de concordância (4,88) quanto ao entendimento sobre o documento mais relevante apresentado pelo sistema. A outra afirmativa, que negava esta primeira, obteve um grau de discordância de 1,25. Desta forma, constata-se que a interface proporcionou a percepção dos documentos mais relevantes utilizando o aspecto de proximidade. A Figura 52 ilustra outro mapa de calor para apresentar a interação dos usuários com os documentos apresentados após a busca.

Figura 52: Mapa de calor da interação dos usuários com os documentos recuperados



Fonte: Elaborado pelo autor

A maioria das interações foram realizadas no documento mais próximo, e também se nota a interação com os outros documentos subsequentes, em menor escala. Houve também interações na área de apresentação do documento e na área de busca, tendo em vista a curiosidade dos participantes em relação à realização de outras pesquisas.

As afirmativas 3 e 7, dispostas no Quadro 13, abordava a tarefa 3 que consistia em analisar detalhes do documento. Ocorreu um alto índice de concordância (4,88), visto que todos os usuários conseguiram completar tal tarefa, a partir da seleção com o mouse, do documento escolhido. Fato também observado na tarefa 4 (acessar o documento na íntegra). Neste caso, a afirmativa consistiu em indicar uma negativa sobre a realização da tarefa. Os usuários indicaram discordância sobre a questão, visto que realizaram com sucesso a atividade.

A Tarefa 5 já foi indicada como o ponto frágil da interface quando examinado o sucesso da realização das atividades. A opinião dos usuários alcançou um nível de concordância mediano (3,63) quanto a perceber a representatividade das palavras em relação aos documentos. Conforme assinalado anteriormente, verifica-se que um caminho a ser investigado seria o uso de ferramentas de apoio como tesouros e ontologias, visto que poderiam fornecer outras significações, além das relações dos termos com os documentos ou outras representações.

As outras situações consideradas no questionário dizem respeito ao uso geral da interface em relação a questões de complexidade e facilidade de uso. Os usuários consideram a ferramenta adequada, assim como uma possível utilização em situações reais de busca. Os usuários indicaram que poderiam utilizar a ferramenta sem necessitar de apoio, além de conseguirem encontrar com facilidade os recursos. O ponto a ser discutido diz respeito à confiança na utilização da ferramenta em que o nível de concordância foi de 3,38 (na escala de 1 a 5). Em relação a este ponto, os usuários apontaram nas sugestões o fato de não estarem familiarizados com este tipo de disposição, e as dúvidas em relação principalmente à questão que relacionava as palavras que representam os documentos, confirmando a necessidade de melhorar a abordagem sobre tal situação.

Por fim, para analisar também as questões de usabilidade, foi utilizada uma escala de diferencial semântico. O objetivo foi variar o formato das questões no que diz respeito a essas questões. A pontuação variou de "-3" a "3", onde os aspectos considerados positivos (simples, útil e intuitiva) foram considerados com valor máximo "3" e suas oposições (complexo, inútil e não intuitiva) indicavam números negativos, com o máximo de

"-3".

Quadro 14: Escala semântica de usabilidade

Situação avaliada	Nível (-3 a 3)
Simples/Complexo	2,38
Útil/Inútil	2,63
Intuitiva/Não intuitiva	2,38

Fonte: Elaborado pelo autor

O questionário e as observações apresentaram um bom índice de aprovação da ferramenta, embora tenham ocorrido indicações pertinentes a situações específicas que merecem atenção especial e continuação das investigações. A escala semântica apresentou uma avaliação bastante positiva nas questões relativas às impressões da interface em relação à simplicidade, utilidade e facilidade de interação.

A utilização de ferramentas diversificadas de coleta de dados se apresentou como um formato adequado, pois as impressões dos usuários sobre o uso da ferramenta foram confirmadas a cada observação coletada. A especificação das etapas a partir do framework DECIDE teve como objetivo organizar a realização das tarefas, assim como a forma de abordagem que utilizou o ensaio de interação.

5.6 SISTEMATIZAÇÃO E DISCUSSÃO

Neste capítulo foi apresentada interface desenvolvida nesta pesquisa, que utiliza como base a estrutura formal proposta no modelo vetorial. Inicialmente abordou-se o projeto do *software* através da UML, elencando fatores estáticos e dinâmicos por meio dos diagramas de classe e casos de uso. Essas especificações foram primordiais para o andamento do projeto. Em seguida foram descritas as tecnologias utilizadas na implementação, iniciando pela explanação do funcionamento da arquitetura cliente/servidor, e explicando o motor de busca Solr que tem como base o modelo vetorial.

Para a interface foram descritos o HTML, CSS, o JavaScript e o PHP que se mostraram adequados ao desenvolvimento. Utilizando as telas do protótipo, foi possível ilustrar suas características visuais e descrever aspectos técnicos que envolveram as tecnologias utilizadas e descritas anteriormente.

Buscou-se também realizar a conexão entre o desenvolvimento da interface e o aporte teórico das áreas de RI, semiótica e IHC. Os conceitos provindos dessas áreas serviram como base para a inserção e disposição dos elementos visuais, além de fornecer uma base teórica sólida. Além disso, os testes de usabilidade foram essenciais para aferir pontos

positivos e fragilidades da proposta. Por ser um protótipo, o sistema ainda necessita de melhorias em sua interface, mas em especial é essencial a realização de mais testes e aprofundamento em alguns pontos, a fim de aprimorar a interface proposta.

6 CONSIDERAÇÕES FINAIS

Investigar a correlação entre a recuperação de informação, a IHC e a semiótica, mostrou-se uma abordagem relevante na pesquisa por recursos que apoiem o progresso no processo de busca informacional, em particular na área de interfaces. Torna-se válido relembrar a questão que norteou este trabalho: "como combinar recursos visuais e textuais à interfaces de sistemas de recuperação de informação para contribuir com o processo de percepção dos usuários em relação aos resultados das buscas". A partir dessa conjectura e das bases da recuperação de informação, foram discutidos os recursos oriundos da IHC e da semiótica que pudessem apoiar a concepção de uma interface a ser utilizada no processo de RI.

Dentro da RI foram analisados os principais conceitos que envolvem o processo de busca por informação, bem como foram identificados os modelos clássicos estudados pela área, em especial o modelo vetorial adotado no desenvolvimento da proposta. Destaca-se que as técnicas do modelo vetorial são adequadas para a representação de documentos, pois devido à diversidade de formatos, esse modelo proporciona um esquema simples, rápido e transparente de representação e ranqueamento de documentos. Dessa maneira, pode-se afirmar que a representação de documentos através do modelo vetorial fornece subsídios necessários para que a descrição e busca desses recursos seja realizada de forma satisfatória. Tanto a pontuação (*score*) como a representação baseada em termos extraídos do próprio documento, são essenciais para demonstrar os critérios de eleição dos documentos recuperados.

Além dos conceitos básicos da RI, outro aspecto apresentado diz respeito ao comportamento informacional dos usuários. O tema apresenta fatores subjetivos e complexos como necessidade de informação, comportamento de busca e relevância, entretanto também evidenciou estudos que sugerem o desenvolvimento de novos estilos de interface como um dos caminhos a serem seguidos na melhoria do processo de busca por informação.

Devido aos objetivos tratarem de questões envolvendo interfaces, os aspectos derivados da IHC exerceram um papel imprescindível na elaboração da proposta, visto que a análise dos estilos de interação foi capaz de elucidar as formas de acesso e manipulação dos elementos a serem dispostos na interface. Além disso, a técnica de visualização RadViz se relaciona oportunamente com os vetores numéricos resultantes da busca através do modelo vetorial.

Durante o desenvolvimento desta tese foram elencados sistemas que

propõem a utilização de elementos visuais na apresentação dos resultados, cada qual com suas especificidades, como a utilização de clusters, mapas e cores, por exemplo. Entretanto, não foram encontradas referências a mecanismos de busca na web em relação a iniciativas comerciais bem-sucedidas e que estão em uso.

A aplicação de conceitos da semiótica aos sistemas de recuperação de informação, objetiva a compreensão de alguns fenômenos de percepção que dificultam a interpretação correta das informações contidas nos documentos. Nesta visão, os fenômenos se apresentam por meio de signos que são percebidos e interpretados e podem influenciar o processo de percepção durante a busca por informação. Os tipos de mediação (icônica, indicial e simbólica) se manifestam na interação humano computador com a finalidade de sustentar a questão cognitiva na visualização de informações.

O Capítulo 5 demonstra a materialização da interface. Iniciou-se a explanação descrevendo os requisitos através de diagramas fornecidos pela UML. Em seguida, foram investigadas as tecnologias que proporcionam a implementação prática das ideias. De um lado, o Solr como base da indexação e do motor de busca e, do outro, as linguagens para a produção da interface (PHP, HTML, CSS e Javascript). Verificou-se também que essas tecnologias disponíveis para implementação da interface fornecem um modelo que oferece condições de implantação dos recursos pretendidos. Outro ponto importante diz respeito à portabilidade do SRI proposto, ou seja, pode ser executado independentemente do sistema operacional ou navegador, visto que todas as tecnologias utilizadas são compatíveis com os principais navegadores de internet utilizados pelos usuários.

Foram utilizadas imagens e códigos da interface para ilustrar a estrutura de interação e disposição dos elementos, além de explicitar aspectos técnicos do produto. Por fim, realizou-se a discussão dos elementos da interface para evidenciar o aporte tanto da RI como da IHC e da semiótica. Também foi realizada uma avaliação de usabilidade que indicou quais formas alternativas em sistemas de RI podem ser desenvolvidas. Foram confrontados os resultados com fundamentos semióticos discutidos como a questão dos aspectos dos signos, o perfil dos internautas e os aspectos perceptivos e cognitivos. Obteve-se uma sinalização positiva nos experimentos por parte dos usuários, mas também foi detectada uma fragilidade com relação à forma como se realizou a representatividade dos documentos.

Pretende-se que as contribuições desta tese possam trazer benefícios para pesquisadores e comunidade que necessitem recuperar informação nas mais variadas plataformas. Dessa forma, uma das principais contribuições deste trabalho diz respeito à

incorporação da visualização diagramática dos documentos resultantes em uma interface de busca por informação. A ideia de utilização de signos e disposição dos elementos tem como base a semiótica e tornou-se possível através da clareza e simplicidade do modelo vetorial. A viabilidade técnica pôde ser determinada a partir da Engenharia de Software, em especial da UML, em conjunto com as referências da área de desenvolvimento de *software*.

Com o avanço no desenvolvimento da interface pretende-se realizar testes comparativos com as ferramentas tradicionais como por exemplo o Google, visto que o protótipo ainda não está totalmente funcional e ainda necessita de ajustes e possivelmente mais estudos para agregar elementos como imagens, sons ou outros artifícios que possam melhorar o processo cognitivo. É possível vislumbrar como trabalho futuro a realização desses experimentos, ou seja, análises que permitam contrapor a utilização de elementos visuais e as interfaces atuais que utilizam listas textuais. Outro estudo futuro foi delineado a partir da fragilidade detectada na interface. Propõe-se, assim, explorar as linguagens documentárias no que tange à representatividade dos documentos a partir de palavras-chave. Nessa perspectiva, supõe-se que alguns instrumentos discutidos nesse contexto, como ontologias e tesouros, sejam úteis na resolução de problemas de representatividade dos documentos.

Vale ressaltar, ainda, a tese peirceana do pensamento diagramático, pois dá sustentação a sistemas de recuperação mais amigáveis, não por ser mais fácil, mas por ser mais efetivo e respeitar a estrutura do pensamento do usuário. Enfim, a partir da tese proposta, acredita-se que é possível relacionar ao desenvolvimento de interfaces de RI os recursos visuais e de interação fundamentados na semiótica e viabilizados pela IHC. Inicialmente essas ideias não visam substituir os modelos atuais de lista textual, mas apresentar uma alternativa de interface viável e exequível que possua referências consistentes na literatura.

REFERÊNCIAS

ADAMI, L. J. **Desenvolvimento de técnicas anycast na camada de aplicação para provisão de qualidade de serviço em computação em nuvem**. 86 f. Dissertação (Mestrado em Ciências da Computação e Matemática Computacional) – Universidade de São Paulo, São Carlos, 2015.

AHN, J.; BRUSILOVSKY, P. **Adaptive visualization for exploratory information retrieval**. Elsevier, 2013. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0306457313000137>>. Acesso em: 20 jul. 2017.

ALEXANDRE, D. S.; TAVARES, J. M. R. S. **Factores da percepção visual humana na visualização de dados**. APMTAC, Portugal, 2007. Disponível em: <https://www.researchgate.net/publication/37649697_Factores_da_Percepcao_Visual_Humana_na_Visualizacao_de_Dados>. Acesso em: 10 nov. 2015.

ALMEIDA, C.C. **Peirce e a organização da informação**: contribuições teóricas da semiótica e do pragmatismo. Marília, 2009. 416 f. Tese (Doutorado em Ciência da Informação) - PPGCI- Faculdade de Filosofia e Ciências, UNESP, Marília, 2009.

ANDERSEN, P. B. What semiotics can and cannot do for HCI. **Knowledge-Based Systems**, v. 14, n. 8, p. 419-424, 2001.

ANDRADE, S. M. et al. Construção e evidências psicométricas de uma escala de avaliação da percepção visual. **Psicologia: Reflexão e Crítica**, Porto Alegre, v. 25, p. 21-29, 2012.

ARTERO, A. O. **Estratégias para apoiar a detecção de estruturas em visualizações multidimensionais perceptualmente sobrecarregadas**. 144 f. Tese (Doutorado em Ciências Matemáticas e de Computação) – Universidade de São Paulo, São Carlos, 2005.

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Recuperação de Informação**: Conceitos e Tecnologias das Máquinas de Busca. 2. ed. Porto Alegre: Bookman, 2013.

BARAT, A. H. The relationship between human perception and knowledge organization. In: **La dimensió humana de l'organització del coneixement**. Facultat de Biblioteconomia i Documentació, p. 286-295, 2005.

BARBOSA, S. D. J.; SILVA, B. S. DA. **Interação Humano-Computador**. Rio de Janeiro: Elsevier, 2011.

BARROS, C. M.; CAFÉ, L. M. A. Estudos da semiótica na Ciência da Informação: relatos de interdisciplinaridades. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 17, n. 3, p. 18-33, jul./set. 2012.

BARROS, L. A. Estruturas morfosintáticas e léxico-semânticas dos termos da Dermatologia. In: ISQUERDO, A. N.; ALVES, I. M. (org.) **As Ciências do Léxico** . Vol. III. Campo Grande: UFMS/São Paulo: Humanitas, p. 397-407, 2007.

BEHESHTI, J. et al. Tracking Middle School Students' Information Behavior Via Kuhlthau's ISP Model: Temporality. **Journal of the American Society for Information Science and Technology**, v.66, p. 943–960, 2014.

BELKIN, N. J.; ODDY, R. N.; BROOKS, H. M. ASK for Information Retrieval: Background and Theory. **Journal of Documentation**, v. 38, n.2, , p.61-71, 1982.

BORGES, G. S. B. **Indexação automática de documentos textuais: proposta de critérios essenciais**. 111 f. Dissertação (Mestrado em Ciência da Informação) – Universidade Federal de Minas Gerais, Belo Horizonte, 2009.

BORLUND, P. The Concept of Relevance in IR. **Journal of the American Society for Information Science and Technology**, v.54, p.913-925, 2003.

BRANSKI, R. M. Recuperação de informações na Web. **Perspectivas em Ciência da Informação**, v. 9, n. 1, p. 70–87, 17 ago. 2004.

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

BUCKLAND, M. What is a digital document? **Document Numérique**, v. 2, n. 2, p. 221–230, 1998.

BUSH, V. **As we may think**. 1945. Disponível em: <<http://www.theatlantic.com/unbound/flashbks/computer/bushf.htm>>. Acesso em: 28 mai. 2018.

CAPES. Coordenação de Aperfeiçoamento de Pessoal de Nível Superior. **Qualis**. 2017. Disponível em: <<http://www.capes.gov.br/acessoainformacao/perguntas-frequentes/avaliacao-da-pos-graduacao/7422-qualis>>. Acesso em: 15 mar. 2018.

CASE, D. O. **Looking for information: a survey of research on information seeking, needs, and behavior**. Amsterdam: Academic Press, 2012. 350 p.

CHAUBEY, R.; SURESH, J. K. Integration vs. development: an engineering approach to building Web applications. In: **Proceedings 2001 Australian Software Engineering Conference**. IEEE, New Canberra, ACT, Australia, 2001. p. 171-181.

CHAUI, M. **Convite à filosofia**. 7.ed. SÃO PAULO: Ática, 2000.

CHOO, C. W.; DETLOR, B.; TURNBULL, D. A behavioral model of information seeking on the Web: preliminary results of a study of how managers and IT specialist use the web. In: ASIS Annual Meeting. **Anais...** Toronto: Fis, 1998.

CHOWDHURY, G. Usuários da Recuperação da Informação. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 5, n.2, p. 4-33, set. 2014/fev. 2015.

CINTRA, A. M. M. Elementos de linguística para estudos de indexação. **Ciência da Informação**, Brasília, v. 12, n. 1, p. 5-22. 1983.

COLLARES, L.; TAVARES, F. T.; GOOCH, A.; TZANETAKIS, G. Personalizing self-organizing music spaces with anchors. **Multimedia Tools and Applications**, v. 77, n. 5, p. 5525–5545, 2018.

CORREIA, J. S. B. C. **Indexação de Documentos Clínicos**. 65 f. Dissertação (Mestrado Integrado em Engenharia Informática e Computação) – Universidade do Porto, Porto, 2016.

CORREIA, M. F. P. V. **Técnicas computacionais na percepção visual do movimento**. 2001. 211 f. Tese (Doutorado em Engenharia Eletrotécnica) – Universidade Fernando Pessoa, Porto, 2001.

CRUZ, C. H. B. Vannevar Bush: uma apresentação. **Revista Latinoamericana de Psicopatologia Fundamental**, v. 14, n. 1, p. 11–13, mar. 2011.

DALE, N.; LEWIS, J. **Ciência da Computação**. 4. ed. Rio de Janeiro: LTC.

DANTAS, C. DA S.; SILVA, T. V. G. DA; SOUZA, A. C. B. Processo de recuperação da informação: barreiras encontradas pelos usuários. **Biblionline**, v. 9, n. 1, p. 16–26, 2013.

DERVIN, B. An overview of sense-making research: concepts, methods, and results to date. In: International Communication Association Annual Meeting. **Anais...** Dallas, Texas, 1983.

DIAS, M. P.; CARVALHO, J. O. F. A Visualização da Informação e a sua contribuição para a Ciência da Informação. **DataGramZero**, Rio de Janeiro, v. 8, n.5, out. 2007. Disponível em: <http://www.dgz.org.br/out07/Art_02.htm>. Acesso em: 23 abr. 2018.

DIAS, P. **A abordagem da comunicação multidimensional na concepção e desenvolvimento de interfaces hipermedia**. 1994. Disponível em: <http://www.c5.cl/ieinvestiga/actas/ribie94/II_30_40.html>. Acesso em: 30 mai. 2018.

ELLIS, D. A behavioural approach to information retrieval system design. **Journal of Documentation**, v. 45, n. 3, p. 171–212, 12 mar. 1989.

FARIAS, P. L. O conceito de diagrama na semiótica de Charles S. Peirce. **Tríades em Revista**, v. 1, n. 1, 2012.

FERNEDA, E. **Ontologia como recurso de padronização terminológica em um Sistema de Recuperação de Informação**. 2013. 96 f. Relatório de Pesquisa (Pós-Doutorado em Ciência da Informação) – Universidade Federal da Paraíba, João Pessoa, 2013.

FERNEDA, E. **Recuperação de Informação: Análise sobre a contribuição da Ciência da Computação para a Ciência da Informação**. 137 f. Tese (Doutorado em Ciências da Comunicação) – Universidade de São Paulo, São Paulo, 2003.

FRANCO, J. R.; BORGES, P. M. O conceito de diagrama em Peirce: uma leitura semiótica para além da gramática especulativa. **Cognitio-Estudos: Revista Eletrônica de Filosofia**, São Paulo, v. 14, n. 1, p. 45-54. 2017.

FRANKENTHAL, E. S. O pragmatismo na abdução e na percepção. **Cognitio-Estudos: Revista Eletrônica de Filosofia**, São Paulo, n. 1, p. 1-9. 2004.

FREITAS, C. M. D. S. et al. Introdução à visualização de informações. **Revista de informática teórica e aplicada**, v. 8, n. 2, p. 143–158, 2001.

FURTADO, R. L.; ALCARA, A. R. Modelos de comportamento informacional: uma análise de suas características. 4º Seminário Científico Arquivologia e Biblioteconomia - do outro lado da informação. **Anais...** Marília: 2015.

GAMBARATO, R. R. Peircean semiotics in the context of design praxis: Abduction and perception in dialogue. **Sign Systems Studies**, v. 41, n. 4, p. 424–432, 2013.

GIL, A. C. **Métodos e técnicas de pesquisa social**. 6. ed. São Paulo: Atlas, 2008.

GIORDANO, R. B.; BIOLCHINI, J. C. A. Busca e recuperação da informação científica na web: comportamento informacional de profissionais da informação. **InCID: R. Ci. Inf. e Doc.**, Ribeirão Preto, v. 3, n.1, p. 125-145, jan./jun. 2012.

GRANJA, C. E. S. C. **Música, conhecimento e educação**: harmonizando os saberes na escola. 2005. 147 f. Dissertação (Mestrado em Educação) - Faculdade de Educação da Universidade de São Paulo, São Paulo, 2005.

GUEDES, G. T. A. **UML 2**: uma abordagem prática. 2. ed. São Paulo, SP: Novatec, 2011.

HARTSON, H. R. Human-computer interaction: Interdisciplinary roots and trends. **The Journal of Systems and Software**, v. 43, n. 1, p. 103-118. 1998.

HAUSMAN, C. R. A Semiótica de Peirce aplicada à percepção: O papel dos objetos dinâmicos e dos perceptos na interpretação perceptiva. **Cognitio**, São Paulo, v. 7, n. 2, p. 231-246, jul./dez. 2006.

HOBBS V.; PFITZNER, D.; POWERS, D. **A Survey of Information Retrieval Interfaces**. 2002. Disponível em: <https://www.academia.edu/18544734/A_Survey_of_Information_Retrieval_Interfaces>. Acesso em 21/08/2017.

HOEBER, O. Information Visualization for Interactive Information Retrieval. In: Chiir '18:

Conference on Human Information Interaction and Retrieval. **Anais...** New Brunswick, New Jersey, USA: 2018.

HOFFMAN, P. DNA Visual And Analytic Data Mining. In: **Proceedings of the 8th IEEE Visualization 97 Conference**. IEEE, AZ, USA, p. 437-441, 1997.

HOFFMANN, M. H. G. Cognição e Pensamento Diagramático. In: QUEIROZ, J.; MORAES, L. (Org.). **A lógica de diagramas de Charles Sanders Peirce**: implicações em ciência cognitiva, lógica e semiótica. Juiz de Fora: Editora UFJF, p. 101-133, 2013.

HUANG, L. **A survey on Web information retrieval technologies**. New York, University of New York, 1999. Disponível em: <http://www.iicm.tugraz.at/thesis/cguetl_diss/literatur/Kapitel02/References/Huang_2000/336617.html>. Acesso em: 09 dez. 2017.

HUANG, S. A semiotic view of information: Semiotics as a foundation of LIS research in information behavior. In: **A semiotic view of information: Semiotics as a foundation of LIS research in information behavior**. ASIST, 2007. p. 1-17.

IBRI, I. A. The heuristic exclusivity of abduction in Peirce's philosophy. In: LEO, Rossella F.; MARIETTI, Susanna (ed.). **Semiotics and Philosophy in C. S. Peirce**. Cambridge: Cambridge Scholars Press, 2006.

INÁCIO, L. **Diagramas lógicos**: a importância filosófica dos diagramas na lógica. 2006. Disponível em: <https://www.academia.edu/1225694/Diagramas_Lógicos>. Acesso em: 30 jun. 2018.

ISLAM, M. N. A systematic literature review of semiotics perception in user interfaces. **Journal of Systems and Information Technology**, v. 15, p. 45-77, 2013.

JANSEN, B. J.; SPINK, A.; SARACEVIC, T. Real life, real users, and real needs: a study and analysis of user queries on the web. **Information Processing & Management**, v. 36, n. 2, p. 207-227, 1 mar. 2000.

JORGE, A. M. G.; REZENDE, D. B.; WARTHA, E. J. Visualização, semiótica e teoria da percepção. **Triade**, Sorocaba, SP, v. 1, n. 1, p. 149-166, jun. 2013.

KHAN, M.; KHAN, S. S. Data and Information Visualization Methods, and Interactive Mechanisms: A Survey. **International Journal of Computer Applications**, v. 34, n. 1, p. 1-14, 2011.

KURAMOTO, H. Uma abordagem alternativa para o tratamento e a recuperação de informação textual : os sintagmas nominais. **Ciência da Informação**, v. 25, n. 2, p. 1-18, 1995.

LANCASTER, F. W. **Indexação e Resumos**: teoria e prática. 2. ed. Brasília, DF: Briquet de Lemos, 2004.

LARA, M. L. G. Algumas contribuições da semiologia e da semiótica para a análise das linguagens documentárias. **Ciência da Informação**, Brasília, v. 22, n. 3, p. 223-226, set./dez. 1993.

LARA, M. L. G. É possível falar em signo e semiose documentária? **Encontros Bibli**, Florianópolis, 2º número esp., 2º sem. 2006.

LEITÃO, C. F.; SILVEIRA, M. S.; SOUZA, C. S. Uma Introdução à Engenharia Semiótica: Conceitos e Métodos. In: 12th Brazilian Symposium on Human Factors in Computing Systems. **Anais...** Porto Alegre: 2013.

LIMA, G. A. B. Interfaces entre ciência da informação e a ciência cognitiva. **Ciência da Informação**, v. 32, n. 1, p. 77-87, 2003.

LIU, T.-Y. Learning to Rank for Information Retrieval. **Foundations and Trends® in Information Retrieval**, v. 3, n. 3, p. 225-331, 2009.

LOOMIS, J. M. Percepção visual do espaço: fenomenologia e função. **Arquivos Brasileiros de Oftalmologia**. São Paulo, SP, v. 66, 2003.

LOSEE, R. M. Term dependence: A basis for Luhn and Zipf models. **Journal of the American Society for Information Science and Technology**, v. 52, n. 12, p. 1019-1025, 2001.

LUCENE. **Apache Lucene**. 2016. Disponível em: < <http://lucene.apache.org/>> Acesso em: 15 fev. 2018.

MAGNANI, L. BARDONE, E. Abduction and Web Interface Design. In: MUELLER, GHAOUI, C. (Org.). **Encyclopedia of Human Computer Interaction**. Liverpool: IGI, p.1-7, 2006.

MAIMONE, G. D.; SILVEIRA, N. C.; TÁLAMO, M. DE F. G. M. Reflexões acerca das relações entre representação temática e descritiva. **Informação & Sociedade Estudos**, v. 21, n. 1, p. 27-35, 2011.

MARCHIONINI, G. User Interfaces for Information Retrieval on the WWW. In: 11th Conference on Professional Information Resources. **Anais...** Chapel Hill, USA: 2005.

MARCONDES FILHO, C. Esquecer Peirce? Dificuldades de uma teoria da comunicação que se apoia no modelo lógico e na religião. **Galaxia**, v. 1, n. 25, p. 38-51, 2013.

MARCOS, M. C. **Visual elements in search and information retrieval systems**. 2005. Disponível em: <https://www.upf.edu/hipertextnet/en/numero-3/busqueda_ri.html>. Acesso em: 25 jun. 2018.

MARTÍNEZ-SILVEIRA, M.; ODDONE, N. Necessidades e comportamento informacional: conceituação e modelos. **Ciência da Informação**, v. 36, n. 2, p. 118-127, 2007.

MATSUNAGA, C. Entre ver e interpretar: o papel fundamental do design. In: 7º Congresso Internacional de Design da Informação, 2., 2015. **Anais...** 2015. p. 797-808. Disponível em: <<http://www.proceedings.blucher.com.br/article-details/entre-ver-e-interpretar-o-papel-fundamental-do-design-20260>>. Acesso em 15 nov. 2015.

MILLETO, E. M.; BERTAGNOLLI, S. C. **Desenvolvimento de Software II: Introdução ao desenvolvimento Web com HTML, CSS, JavaScript e PHP**. Porto Alegre: Bookman, 2014.

MONTEIRO, S. D. As múltiplas sintaxes dos mecanismos de busca no ciberespaço. **Informação e Informação**, Londrina, v. 14, n. esp, p. 68-102. 2009.

MOOERS, C. Zetocoding applied to mechanical organization of knowledge. **American Documentation**, v. 2, n. 1, p. 20–32, 1951.

MORAES, J. A.; PEREIRA, P. H. A. O.; PANTALEÃO, N. C. A. Percepção direta: contribuições do pragmatismo peirciano para a abordagem ecológica. **Revista Eletrônica de Filosofia**, São Paulo, v. 9, p. 195-204, 2012.

MOURA, M. A. Ciência da Informação e Semiótica: Conexão de saberes. **Encontros Bibli: Revista Eletrônica de Biblioteconomia e Ciência da Informação**, Florianópolis, n. 2, p. 1-17. 2006.

MOURA, M. A.; SILVA, A. P.; AMORIM, V. R. A concepção e o uso das linguagens de indexação face às contribuições da semiótica e da semiologia. **Informação & Sociedade: Estudos**, v. 12, n. 1, p. 217-251. 2002.

MUCELIN, C. A.; BELLINI, M. Estudo da percepção em ecossistema urbano: uma contribuição para a educação, planejamento e gestão ambiental. **Revista Brasileira de Estudos Pedagógicos**, Brasília, v. 89, n. 221, p. 119-144, jan./abr. 2008.

MÜLLER, H.M.; VAN AUKEN, K. M.; LI, Y.; STERNBERG, P. W. Textpresso Central: a customizable platform for searching, text mining, viewing, and curating biomedical literature. **BMC Bioinformatics**, v. 19, n. 1, p. 1–16, 2018.

NASCIMENTO, A. M. N.; AMARAL, S. A. **Avaliação de Usabilidade na Internet**. Brasília: Thesaurus, 2010.

NASCIMENTO, H. A. DO; FERREIRA, C. B. R. Visualização de informações: uma abordagem prática. In: XXV Congresso da Sociedade Brasileira de Computação. **Anais...** São Leopoldo: 2005.

NETTO, A. A. O. **IHC e a Engenharia Pedagógica**. Florianópolis: Visual Books, 2010.

NIEDERAUER, J. **Desenvolvendo Websites com PHP**. 2. ed. São Paulo: Novatec, 2011.

NIELSEN, J. **Usability Engineering**. Mountain View, California: Morgan Kaufmann, 1993.

NIELSEN, J. **Why You Only Need to Test with 5 Users**. 2000. Disponível em: <<https://www.nngroup.com/articles/why-you-only-need-to-test-with-5-users/>>. Acesso em: 12 dez. 2018.

NÓBREGA, T. P. Corpo, percepção e conhecimento em Merleau-Ponty. **Estudos de Psicologia**. v. 13 (2), p. 141-148, 2008.

NORMAN, D. A. Cognitive Engineering. In: NORMAN, D.; DRAPER, S. (Org.). **User Centered System Design**. Hillsdale, NJ: Lawrence Erlbaum, p. 31–61, 1986.

NÖTH, W. **Panorama da semiótica de Platão a Peirce**. 4. ed. São Paulo: Annablume, 2003.

NÖTH, W. **Semiótica do Século XX**. São Paulo: Annablume, 1996.

NOVELLINO, M. S. F. Instrumentos e metodologias de representação da informação. **Informação & Informação**, v. 1, n. 2, p. 37, 15 dez. 1996.

OLIVEIRA, I. A. Interface de Usuário: A Interação Homem-Computador Através dos Tempos. **Olhar Científico**, v. 1, n. 2, p. 178–184, 17 maio 2010.

ORTEGA, C. D.; LARA, M. L. G. DE. A noção de documento: de Otlet aos dias de hoje. **DataGramZero - Revista de Ciência da Informação**, v. 11, n. 2, p. 1–22, 2010.

PAJÍC, D. Browse to search, visualize to explore: Who needs an alternative information retrieving model? **Computers in Human Behavior**, v. 39, p. 145–153, 2014.

PAVÃO, C. M. G. **Comportamento de busca e recuperação da informação em serviços de descoberta em rede no contexto acadêmico**. 218 f. Tese (Doutorado em Comunicação e Informação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2014.

PEIRCE, C. S. **Semiótica**. São Paulo: Perspectiva, 2005.

PEIRCE, C. S. **The electronic edition of the collected papers of Charles Sanders Peirce**. Vols. I-VI. In: Hartshorne C, Weiss P, eds. Cambridge: Harvard University Press; 1931-1935, 1994.

PEREIRA, L. A. DE M. **Análise e Modelagem de Sistemas com a UML**. 1. ed. Rio de Janeiro: CIP-Brasil, 2011.

PITANGUEIRA, A. M. S. **Caracterização dos fatores de relevância do usuário no processo de recuperação de informação: estudo em ambiente de pesquisa em pós-graduação**. Dissertação (Mestrado em Administração) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2004.

POTTKER, L. M. V. **Arquitetura para recuperação de objetos de aprendizagem: uma abordagem baseada em agentes inteligentes e relevance feedback**. 204 f. Tese (Doutorado

em Ciência da Informação) – Universidade Estadual Paulista, Marília, 2017.

PRATES, R. O.; BARBOSA, S. D. J. Introdução à Teoria e Prática da Interação Humano-Computador fundamentada na Engenharia Semiótica. In: Jornadas de Atualização em Informática. **Anais...** 2007.

PREECE, J. et al. **Human-computer interaction**. Addison-Wesley Pub. Co, 1994.

PREECE, J.; ROGERS, Y.; SHARP, H. **Design de Interação**. Porto Alegre: Bookman, 2005.

PRESSMAN, R. S. **Engenharia de Software - Uma Abordagem Profissional**. 7 ed. McGraw Hill, 2011.

PRESSMAN, R. S. **Engenharia de Software**. 6 ed. São Paulo: McGraw Hill, 2006.

PRODANOV, C. C.; FREITAS, E. C. **Metodologia do trabalho científico: Métodos e Técnicas da Pesquisa e do Trabalho Acadêmico**. 2. ed. Novo Hamburgo: Feevale, 2013.

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, 2013.

REGO, J. M. R.; PESSOA, G. H. R.; GALA, A. C. O. S. A Percepção em três registros: Russell, Peirce e Caeiro. **Revista Eletrônica de Filosofia**, São Paulo, v. 12, p. 98-116, 2015.

RIJSBERGEN, C. J. **Information retrieval**. London: Butterworths, 1979. Disponível em: . Acesso em: 22 mai. 2018.

ROBREDO, J. A indexação automática de textos: o presente já entrou no futuro. **Estudos Avançados em Biblioteconomia e Ciência da Informação**, v. 1, n. 1, p. 235–274, 1982.

ROCHA, H. V. DA; BARANAUSKAS, M. C. C. **Design e Avaliação de Interfaces Humano-Computador**. Campinas: Unicamp, 2003.

ROCHA, L. B. Fenomenologia, semiótica e geografia da percepção: alternativas para analisar o espaço geográfico. **Revista da Casa da Geografia de Sobral**, Sobral, v. 4/5, p. 67-79, 2002/2003.

RODRIGUES, G. A. **Um simulador para modelos descritos em UML**. 92 f. Dissertação (Mestrado em Engenharia) – Universidade de São Paulo, São Paulo, 2009.

ROLIM, E. A.; CENDÓN, B. V. Modelos teóricos de estudos de usuários na ciência da informação. **DataGramZero - Revista de Ciência da Informação**, 2013. Disponível em: <<http://www.brapci.inf.br/index.php/article/download/50794>>

ROMANINI, A. V. **Semiótica minuta: especulações sobre a gramática dos signos e da comunicação a partir da obra de Charles S. Peirce**. 2006. 246f. Tese (Doutorado em Ciências

da Comunicação) - Escola de Comunicações e Artes da Universidade de São Paulo (USP), São Paulo, 2006.

SALES, R. A representação documental como um encontro de representações. In: I EnReDo – Encontro de Representação Documental, 2017. **Anais...** 2017. Disponível em: <<http://www.telescopium.ufscar.br/index.php/enredo/enredo/paper/viewFile/87/86>>. Acesso em 22 dez. 2018.

SALTON, G.; MCGILL, M. J. **Introduction to Modern Information Retrieval**. New York: Mcgraw-Hill Book Company, 1983.

SANTAELLA, L. **A percepção: uma teoria semiótica**. 2. ed. São Paulo: Experimento, 1998.

SANTAELLA, L. **A teoria geral dos signos: semiose e autogeração**. São Paulo: Ática, 1995.

SANTAELLA, L. **Cultura e artes do pós-humano: da cultura das mídias à cibercultura**. São Paulo: Paulus, 2003.

SANTAELLA, L. **Navegar no ciberespaço: o perfil cognitivo do leitor imersivo**. São Paulo: Editora Paulus, 2004.

SANTAELLA, L. **Matrizes da linguagem do pensamento: sonora, visual, verbal**. São Paulo: Iluminuras, 2005.

SANTAELLA, L. **O que é semiótica**. São Paulo: Editora Brasiliense, 2012a.

SANTAELLA, L. **Percepção: fenomenologia, ecologia, semiótica**. São Paulo: Cengage Learning, 2012b.

SANTAELLA, L. **Semiótica aplicada**. São Paulo: Cengage Learning, 2015.

SANTOS, F. V. DA S.; GASPAR, N. R. Um olhar para formas de apresentação de resultados na web: uma proposta da análise do discurso francesa. In: XXV Congresso Brasileiro de Biblioteconomia, Documentação e Ciência da Informação. **Anais...** Florianópolis, SC, Brasil: 2013.

SANTOS, G. S. **Engenharia de Usabilidade**. 2 ed. Rio de Janeiro: Seses, 2016.

SANTOS, N. A. Tópicos em percepção e processamento visual da forma: acuidade visual versus sensibilidade ao contraste. **Estudos e Pesquisas em Psicologia**. Rio de Janeiro, v. 3, 2003.

SARACEVIC, T. Ciência da informação: origem, evolução e relações. **Perspectivas em Ciência da Informação**, Belo Horizonte, v. 1, n. 1, p. 41-62, jan./jun. 1996.

SARACEVIC, T. Relevance: A review of the literature and a framework for thinking on the notion in information science. Part III: Behavior and effects of relevance. **Journal of the**

American Society for Information Science and Technology, v. 58, n. 13, p. 2126–2144, 2007.

SCHULTZ, Sérgio. A análise gödeliana do conceito de percepção. **Doispontos**, Curitiba, São Carlos, v. 9, p. 135-150, out. 2012.

SCOTTINI, A. **Minidicionário de língua portuguesa**. Blumenau: Todolivro Editora, 2009.

SEDGHI, S.; SHORMEIJ, Z.; TAHAMTAN, I. Exploring the context of visual information seeking. **The Electronic Library**, v. 36, n. 3, p. 445-456, 2018.

SHNEIDERMAN, B. **Designing the User Interface**. 3. ed. Addison Wesley, 1998.

SINGH, R.; HSU, Y.-W.; MOON, N. Multiple perspective interactive search: a paradigm for exploratory search and information retrieval on the web. **Multimedia Tools Applications**, v. 62, n. 2, p. 507–543, 2013.

SIQUEIRA, J. C. A Noção de documento digital: uma abordagem terminológica. **Em Questão**, v. 18, n. 1, p. 125–140, 30 jul. 2012.

SOLR. **Apache Solr**. 2017. Disponível em: < <http://lucene.apache.org/solr/>>. Acesso em: 10 fev. 2018.

SOMMERVILLE, I. **Engenharia de Software**. 8. ed. Rio de Janeiro: Prentice-Hall, 2008.

SONG, M. Visualization in information retrieval: a three-level analysis. **Journal of Information Science**, v. 26, n. 1, p. 3–19, 1 fev. 2000.

SOUSA, Hugo Miguel dos Reis Costa. **Percepção humana na visualização de informação crítica**. 2009. 181 f. Dissertação (Mestrado em Tecnologia Multimédia) – Universidade do Porto, Porto, 2009.

SOUZA, C. S. et al. **Projeto de Interfaces de Usuário: Perspectivas Cognitivas e Semióticas**. 1999. Disponível em: <https://www.dimap.ufrn.br/~jair/piu/JAI_Apostila.pdf>. Acesso em: 30 maio. 2018.

SOUZA, C. S. Semiotic engineering: bringing designers and users together at interaction time. **Interacting with Computers**, v. 17, n. 3, p. 317-341, 2005.

SOUZA, C. S. Semiotics. In: SOEGAARD, M.; DAM, R. F. (Eds.). **The Encyclopedia of Human-Computer Interaction**. The Interaction Design Foundation. p. 1–35, 2012.

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

SOUZA, R. R.; ALVARENGA, L. A Web Semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, p. 132–141, 2004.

TOMASI, S.; SCHUFF, D.; TURETKEN, O. Understanding novelty: how task structure and tool familiarity moderate performance. **Behaviour & Information Technology**, v. 37, n. 4, p. 406-418, 2018.

TORRES, R. A. O. **Reordenação de Documentos Através de Técnicas de Aprendizagem Automática**. 75 f. Dissertação (Mestrado em Engenharia Informática) – Universidade Nova de Lisboa, Lisboa, 2014.

VALDIVIA, A. M. C. **Mapeamento de dados multidimensionais usando árvores filogenéticas**: foco em mapeamento de textos. 2007. 87 f. Dissertação (Mestrado em Ciência da Computação e Matemática Computacional). USP - São Carlos. São Carlos, 2007.

VALIATI, E. R. A. **Avaliação de usabilidade de técnicas de visualização de informações multidimensionais**. 220 f. Tese (Doutorado em Ciência da Computação) – Universidade Federal do Rio Grande do Sul, Porto Alegre, 2008.

VANDI, C. Beyond metaphors and icons: towards a perception-action model for graphic user interfaces. In: 9th World Congress of IASS/AIS, 2007, Helsinki. **Anais...** Disponível em: <http://www2.uiah.fi/sefun/DSIU_papers/DSIU_Vandi_Beyond%20metaphors%20and%20icons.pdf>. Acesso em: 20 jul. 2016.

VAZ, F. R.; CARVALHO, C. L. **Visualização de Informações**. Relatório Técnico. Universidade Federal de Goiás. Goiânia, 2004.

VENÂNCIO, L. S.; NASSIF, M. E. O comportamento de busca de informação sob o enfoque da cognição situada: um estudo empírico qualitativo. **Ciência da Informação**, v. 37, n. 1, p. 95–106, 2008.

VIEIRA, J. M. DE L.; CORREA, R. F. Visualização da informação na construção de interfaces amigáveis para sistemas de recuperação de informação. **Encontros Bibli**, v. 16, n. 32, p. 73–93, 2011.

W3SCHOOLS. **W3schools.com**. 2018. Disponível em: <<https://www.w3schools.com/>>. Acesso em: 05 fev. 2018.

WARE, C. **Information Visualization: Perception for Design**. 2. ed. San Francisco: Morgan Kaufmann, 2004.

WILSON, T. D. Models in information behaviour research. **Journal of Documentation**, v. 55, n. 3, p. 249-270, 1999.

WILSON, T. D. Human information behavior. **Informing Science**, v. 3, n. 2, p. 49–55, 2000.

WILSON, T. D. **Evolution in information behavior modeling**: Wilson's model. In:

FISCHER, K.; ERDELEZ, S.; MCKECHNIE (Eds.). Theories of information behavior. *Information Today*. p. 31–36, 2005.

XAVIER, Raphael Figueiredo. **Análise de métodos de produção de interfaces visuais para recuperação da informação**. Dissertação (Mestrado em Ciência da Informação) – Universidade Estadual Paulista (UNESP), Marília, 2009.

XIE, I.; JOO, S. Transitions in search tactics during the Web-based search process. **Journal of the American Society for Information Science and Technology**, v. 61, n. 11, p. 2188–2205, 1 nov. 2010.

APÊNDICES

APÊNDICE A – Roteiro de teste de usabilidade

CENÁRIO 1 - REALIZAR BUSCA BASEADA EM PALAVRAS-CHAVE

Tarefas

TAREFA 1 – Você precisa realizar uma pesquisa baseada em palavra-chave, assim como faz no Google. Seu objetivo é explorar o termo "Recuperação de Informação. Informe o termo de pesquisa na interface e realize a busca.

Interface 1 – Interface Inicial

Realização da tarefa:

() Sim, com facilidade () Sim, com dificuldade () Não conseguiu realizar a tarefa

TAREFA 2 - Analisar os resultados da busca e indicar qual o documento se apresenta como mais relevante. Qual arquivo corresponde à busca realizada?

Interface 3 – Interface com resultados

Realização da tarefa:

() Sim, com facilidade () Sim, com dificuldade () Não conseguiu realizar a tarefa

TAREFA 3 – Como você faria para saber detalhes sobre o documento de seu interesse?

Interface 4 – Detalhamento dos resultados

Realização da tarefa:

() Sim, com facilidade () Sim, com dificuldade () Não conseguiu realizar a tarefa

TAREFA 4 - Como você faria para acessar o documento mais relevante?

Interface 5 – Acesso ao conteúdo

Realização da tarefa:

() Sim, com facilidade () Sim, com dificuldade () Não conseguiu realizar a tarefa

TAREFA 5 – No detalhamento do arquivo aparecem alguns termos na parte inferior, qual sua percepção sobre esses itens.

Interface 6 – Detalhamento dos resultados

Realização da tarefa:

() Sim, com facilidade () Sim, com dificuldade () Não conseguiu realizar a tarefa

APÊNDICE B – Questionário de usabilidade

Perfil dos Participantes

1. Sexo
 Feminino Masculino
2. Faixa Etária
 De 14 a 23 anos De 24 a 33 anos
 De 34 a 43 anos De 44 a 53 anos
 54 anos ou mais
3. Qual é sua formação?
 Ensino Médio
 Ensino Superior
 Especialização
 Mestrado (Concluído)
 Doutorado (Concluído)
4. Qual é a sua profissão? _____
5. Como você avalia os seus conhecimentos em Informática?
 Ótimo Bom Regular Ruim
6. Com qual frequência você utiliza os serviços de busca na Web (Google, Bibliotecas Digitais, etc.)?
 Diariamente Semanalmente
 Mensalmente Raramente
7. A quanto tempo utiliza esses serviços?
 2 anos ou menos 3 a 5 anos
 6 a 8 anos mais de 8 anos
8. Qual o principal motivo da utilização desses serviços?
 Atividades Profissionais Lazer
 Fins Acadêmicos Compras
- Qual serviço utiliza? _____

Avaliação de Usabilidade

10. Avalie seu grau de discordância/ concordância com as seguintes afirmações sobre a ferramenta testada.

	Discordo totalmente		Neutro		Concordo Totalmente
Fatores	1	2	3	4	5
1. Consegui realizar as buscas e acessar os documentos na interface disponibilizada.					
2. Consegui identificar o documento mais relevante indicado pela interface.					
3. A disposição dos elementos permite acessá-los rapidamente					
4. Os recursos de navegação (ícones e botões) estão todos claros e fáceis de achar.					
5. Não consegui acessar o documento que selecionei.					
6. Eu achei a ferramenta complexa. Precisei pensar muito para completar as tarefas.					
7. Os elementos dispostos na interface proporcionaram acesso às informações dos documentos.					
8. Eu precisaria de apoio de uma pessoa para usar este aplicativo.					
9. Eu me senti muito confiante usando esta ferramenta.					
10. Consegui visualizar as palavras que representam o documento e seu grau de importância.					
11. Não foi possível identificar qual documento era o mais relevante na busca.					

12. Essa ferramenta seria útil nas minhas atividades de pesquisa.						
---	--	--	--	--	--	--

11. Avalie a sua percepção em relação ao uso da interface. Assinale com um "x" quando acredita que o atributo representa sua opinião. Deve ser assinalado apenas um "x" em cada linha

	3	2	1	0	-1	-2	-3	
Simple								Complexo
Útil								Inútil
Intuitiva								Não intuitiva

12. Descreva eventuais dúvidas e sugestões:

APÊNDICE C – Carta de consentimento

Carta de consentimento

Afirmo que desejo participar do plano de teste que está sendo conduzido pelo pesquisador Fabricio Baptista em pesquisa de Doutorado realizada no programa de Pós Graduação em Ciência da Informação da Universidade Estadual Paulista - Campus Marília.

O propósito desta pesquisa é avaliar a usabilidade da interface de busca proposta pela pesquisa. Responderei a algumas questões demográficas e também sobre a experiência de uso da interface.

Todas as informações contidas neste estudo são confidenciais, e meu nome não será identificado em momento algum.

Estou ciente que posso fazer perguntas ou desistir da colaboração em qualquer momento, sem qualquer tipo de penalidade.

Adaptado de (Preece, Rogers e Sharp, 2005)

APÊNDICE D – Tabulação dos Dados

Perfil dos Participantes

Participante	Sexo	Faixa etária	Formação	Profissão	Conhecimentos em Informática	Utilização de sistemas de busca	Experiência em mecanismos de busca	Qual serviço utiliza	Principal Motivo
1	Masculino	De 14 a 23 anos	Ensino Médio	Estudante	Ótimo	Diariamente	3 a 5 anos	Google	Lazer
2	Feminino	De 14 a 23 anos	Ensino Médio	Estudante	Bom	Diariamente	2 anos ou menos	Google	Lazer
3	Feminino	De 34 a 43 anos	Especialização	Pedagogo	Regular	Diariamente	mais de 8 anos	Google	Profissionais
4	Masculino	De 24 a 33 anos	Graduação	Professor	Ótimo	Diariamente	mais de 8 anos	Google	Profissionais
5	Feminino	De 44 a 53 anos	Especialização	Tecnico Adm	Ótimo	Diariamente	mais de 8 anos	Google	Profissionais
6	Feminino	De 14 a 23 anos	Ensino Médio	Estudante	Bom	Semanalmente	2 anos ou menos	Google	Acadêmicos
7	Feminino	De 34 a 43 anos	Mestrado	Tecnico Adm	Regular	Diariamente	mais de 8 anos	Google	Profissionais
8	Masculino	De 14 a 23 anos	Ensino Médio	Estudante	Regular	Diariamente	3 a 5 anos	Google	Lazer

Fonte: Elaborado pelo autor

Execução das tarefas

Participante	Tempo de teste (minutos)	Tempo Questionario (minutos)	Tempo total	Tarefa 1	Tarefa 2	Tarefa 3	Tarefa 4	Tarefa 5
1	2	2	15	Concluiu	Concluiu	Concluiu	Concluiu	Concluiu
2	4	3	20	Concluiu	Concluiu	Concluiu	Concluiu	Parcial
3	3	3	18	Concluiu	Concluiu	Concluiu	Concluiu	Não concluiu
4	3	3	18	Concluiu	Concluiu	Concluiu	Concluiu	Não concluiu
5	2	4	15	Concluiu	Concluiu	Concluiu	Concluiu	Concluiu
6	3	4	20	Concluiu	Concluiu	Concluiu	Concluiu	Parcial
7	3	3	17	Concluiu	Concluiu	Concluiu	Concluiu	Parcial
8	4	3	17	Concluiu	Concluiu	Concluiu	Concluiu	Parcial

Fonte: Elaborado pelo autor

Questionário de usabilidade

N	Questão	Grau de concordância	1	2	3	4	5	6	7	8
1	1. Consegui realizar as buscas e acessar os documentos na interface disponibilizada.	4,63	5	5	4	4	5	5	4	5
2	2. Consegui identificar o documento mais relevante indicado pela interface.	4,88	5	5	5	5	4	5	5	5
11	11. Não foi possível identificar qual documento era o mais relevante na busca.	1,25	2	1	1	2	1	1	1	1
3	3. A disposição dos elementos permite acessá-los rapidamente.	4,88	5	5	5	5	4	5	5	5
7	7. Os elementos dispostos na interface proporcionaram acesso às informações dos documentos.	4,88	5	5	5	5	4	5	5	5
5	5. Não consegui acessar o documento que seleccionei.	1,50	2	1	1	2	2	1	1	2
10	10. Consegui visualizar as palavras que representam o documento e seu grau de importância.	3,63	4	2	3	4	5	4	4	3
4	4. Os recursos de navegação (ícones e botões) estão todos claros e fáceis de achar.	4,50	4	5	4	5	4	4	5	5
6	6. Eu achei a ferramenta complexa. Precisei pensar muito para completar as tarefas.	1,13	2	1	1	1	1	1	1	1
8	8. Eu precisaria de apoio de uma pessoa para usar este aplicativo.	1,88	2	1	2	2	2	2	2	2
9	9. Eu me senti muito confiante usando esta ferramenta.	3,38	3	4	1	1	4	5	4	5
12	12. Essa ferramenta seria útil nas minhas atividades de pesquisa.	4,38	4	4	4	4	5	5	4	5

Fonte: Elaborado pelo autor

Usabilidade através de escala semântica

N	Ques tão	Nível (-3 a 3)	1	2	3	4	5	6	7	8
1	Simple s/Complexo	2,38	2	3	2	3	2	2	3	2
2	Útil/Inútil	2,63	3	3	2	2	3	2	3	3
3	Intuitiva/Não intuitiva	2,38	2	3	2	2	3	2	3	2

Fonte: Elaborado pelo autor