

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMA DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

ANTÔNIO MAURÍCIO DA SILVA PITANGUEIRA

**CARACTERIZAÇÃO DOS FATORES DE RELEVÂNCIA DO USUÁRIO
NO PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO: estudo em
ambiente de pesquisa em pós-graduação**

Porto Alegre, 2004

**UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
ESCOLA DE ADMINISTRAÇÃO
PROGRAMAÇÃO DE PÓS-GRADUAÇÃO EM ADMINISTRAÇÃO**

Antônio Maurício da Silva Pitangueira

**CARACTERIZAÇÃO DOS FATORES DE RELEVÂNCIA DO USUÁRIO
NO PROCESSO DE RECUPERAÇÃO DE INFORMAÇÃO: estudo em
ambiente de pesquisa em pós-graduação**

Dissertação de Mestrado, apresentada ao Programa de Pós-Graduação em Administração da Universidade Federal do Rio Grande do Sul como requisito parcial para a obtenção do título de Mestre em Administração.

Orientador: Prof. Dr. Eduardo Ribas Santos

Porto Alegre, 2004

Dedico este trabalho às
pessoas que mais amo:
minha mãe e meu pai.

AGRADECIMENTOS

Expresso aqui meus sinceros agradecimentos a todos que contribuíram para a realização desta pesquisa, em especial para:

- Meu orientador e grande amigo, pela sua paciência, idéias extremamente relevantes e o apoio incondicional. Espero que eu seja semelhante nas minhas futuras orientações;
- Os professores João Luiz Becker e Denis Borenstein, pela disponibilidade em sugerir e recomendar aspectos interessantes que contribuíram para a efetivação desta pesquisa;
- Lourdes do CEPA e sua equipe, pelo constante suporte na análise estatística;
- Os entrevistados, que possibilitaram a coleta dos dados;
- O CNPQ, pelo suporte financeiro e ajuda na minha formação;
- Os professores do PPGA, que enriqueceram minha formação acadêmica;
- Os antigos amigos e aos novos, por apostarem em mim e ajudaram nos momentos críticos;
- Minha irmã e minha dinda, pelo constante apoio;
- Todos os meus familiares.

RESUMO

Com o desenvolvimento e crescimento das tecnologias de informação e, em especial, a Internet, encontrar informação relevante e útil tornou-se uma tarefa complexa que envolve os aspectos computacionais dos Sistemas de Recuperação de Informação (SRIs) e os aspectos situacionais, psicológicos e cognitivos dos usuários. No entanto, há um desequilíbrio neste cenário pois o enfoque das pesquisas de recuperação de informação (RI) está extremamente voltado para a abordagem computacional. Decorrente deste viés, o aspecto central da recuperação de informação – o julgamento de relevância – tornou-se muito limitado e acarretou em baixo desempenho dos SRIs. Notando esta problemática, esta pesquisa teve como objetivo caracterizar os fatores de relevância empregados pelos usuários durante o processo de RI com o intuito de contribuir para a especificação de requisitos e avaliação daqueles sistemas. Assim, a pesquisa foi composta por duas etapas de avaliação denominadas *ex ante* e *ex post* para verificar a mudança de percepção em torno dos julgamentos de relevância. Para tanto, foi criado um instrumento de pesquisa para verificar os julgamentos de relevância que foram avaliados por 49 respondentes, obtendo um total de 238 objetos informacionais avaliados no momento *ex ante* e destes, 224 foram retornados (*ex post*). Os resultados encontrados mostraram que existe uma significativa mudança de percepção dos usuários nas duas medições, e demonstraram que os fatores de relevância obedecem a um processo decisório bem definido. Observou-se também que a utilidade e os aspectos cognitivos são as manifestações de relevância consideradas mais importantes pelos usuários e as demais servem para ter a noção de rejeição ou não do objeto informacional.

Palavras-chave: recuperação de informação, fatores de relevância, critérios de relevância.

ABSTRACT

With the development and growth of information technologies and, specially, of the Internet, getting excellent and useful information became a complex task that involves the computational aspects of the Information Retrieval Systems (IRS) and the situational, psychological and cognitive aspects of the users. However, there is a disequilibrium in this scenery because the focus of the research on information retrieval (IR) is extremely turned to the computational approach. As a consequence of this bias, the central aspect of the information retrieval - the relevance judgment- became very limited and it caused bad performance of the SRIs. Noticing this problematic, the present research was aimed to characterize the relevance factors employed by the users during the process of IR with the intention to contribute for the specification of requirements and evaluation of those systems. Thus, the research was composed by two stages of evaluation named *ex ante* and *ex post*, to verify the change of perception about the judgments relevance. For in such a way, a research instrument was created to verify the judgments relevance that had been evaluated by 49 respondents, getting a total of 238 informational objects evaluated (*ex ante*) and from that 224 *ex post* evaluations were returned. The results had shown that it exists a significant change in the perception of the users in the two measurements, and had demonstrated that the relevance factors obey to a very precise decision process. It was also observed that the utility and cognitive aspects are the more important relevance manifestations for the users and that the others serve to have the notion of rejection or acceptance the informacional object.

Key Words: information retrieval, relevance factors, relevance criteria.

SUMÁRIO

LISTA DE FIGURAS

LISTA DE QUADROS

LISTA DE TABELAS

1 INTRODUÇÃO	12
2 CONTEXTUALIZAÇÃO DA PESQUISA	14
2.1 O PROCESSO PARA A DEFINIÇÃO DOS FATORES DE RELEVÂNCIA	18
2.1.1 O Sistema de RI.....	18
2.1.2 O Usuário e o Processo	19
2.2 SITUAÇÃO PROBLEMÁTICA	20
2.3 OBJETIVOS	22
OBJETIVO GERAL.....	22
OBJETIVOS ESPECÍFICOS	22
3 RECUPERAÇÃO DE INFORMAÇÃO	24
3.1 PROCESSO E MODELOS DE RECUPERAÇÃO DE INFORMAÇÃO	25
4 RELEVÂNCIA	36
4.1 A EVOLUÇÃO DO CONCEITO DE RELEVÂNCIA	37
4.1.1 Tipos de Relevância	39
4.1.2 Critérios de Relevância.....	42
4.1.3 Regiões de Relevância.....	49
4.1.4 Medidas de Relevância	50
5 MÉTODO DE PESQUISA	52
5.1 CARACTERIZAÇÃO DA PESQUISA	52
5.2 SELEÇÃO DOS PARTICPANTES DA PESQUISA – FASE A	53
5.3 SEÇÃO DE REFERÊNCIA – FASE B	54
5.4 AVALIAÇÃO DE OBJETOS INFORMACIONAIS – FASE C.....	55
5.4.1 Instrumento de Pesquisa	55
5.4.2 Procedimento de Coleta dos Dados	56
5.5 ANÁLISE – FASE D.....	58
6 RESULTADOS DA PESQUISA	60
6.1 ETAPA EXPLORATÓRIA QUALITATIVA	61
6.1.1 Definição das Estratégias de Busca	61
6.1.2 Formulação das Consultas e Seleção Inicial dos OIs	62
6.1.3 Critérios Indicados pelos Respondentes.....	63
6.2 ETAPA EXPLORATÓRIA QUANTITATIVA	65
6.2.1 Teste de Consistência Interna	65
6.2.2 Caracterização da Amostra.....	66
6.2.3 Atividade Acadêmica	66
6.2.4 Finalidade	67
6.2.5 Nível de Conhecimento	67
6.2.6 Bases de Busca Utilizadas.....	68
6.2.7 Distribuição das Respostas.....	68
6.2.8 Análise Univariada.....	69
6.2.9 Análise Multivariada	74

7 CONSIDERAÇÕES FINAIS.....	93
7.1 CONCLUSÕES.....	93
7.2 IMPLICAÇÕES DA PESQUISA.....	95
7.2.1 Estudos sobre Relevância.....	95
7.2.2 Projetos de Sistemas de RI.....	96
7.3 LIMITAÇÕES DA PESQUISA.....	97
7.4 SUGESTÕES PARA PESQUISAS FUTURAS.....	97
REFERÊNCIAS BIBLIOGRÁFICAS.....	98
ANEXOS.....	103

LISTA DE FIGURAS

Figura 1	Contextualização da dissertação	17
Figura 2	Processo sob o ponto de vista do usuário	25
Figura 3	Processo sob o ponto de vista do sistema	26
Figura 4	Modelo tradicional de RI	27
Figura 5	Modelo cognitivo de interação em RI	29
Figura 6	Modelo de interação em RI baseado em eventos	31
Figura 7	Modelo de interação em RI baseado em camadas	32
Figura 8	Efeitos de busca no espaço total da base de dados	36
Figura 9	Fases do método de pesquisa	52
Figura 10	Fatores de relevância para a topicalidade intelectual	77
Figura 11	Fatores de relevância para a estrutura/ formato	80
Figura 12	Fatores de relevância para a manifestação cognitiva	83
Figura 13	Fatores de relevância para a utilidade	86
Figura 14	Fatores de relevância para a manifestação motivacional	88
Figura 15	Visão geral das manifestações de relevância	91

LISTA DE QUADROS

Quadro 1	Componentes e níveis do modelo em camadas	33
Quadro 2	Exemplos de tipos de relevância	39
Quadro 3	Manifestações de relevância	41
Quadro 4	Síntese dos critérios de relevância	43

LISTA DE TABELAS

Tabela 1	Testes de confiabilidade	65
Tabela 2	Composição da amostra	66
Tabela 3	Atividade acadêmica	66
Tabela 4	Finalidade da informação	67
Tabela 5	Nível de conhecimento por atividade acadêmica	67
Tabela 6	Bases de busca utilizadas	68
Tabela 7	Distribuição das respostas de relevância	68
Tabela 8	Comportamento geral dos critérios de topicalidade intelectual	69
Tabela 9	Comportamento geral dos critérios da manifestação de estrutura/ formato ..	70
Tabela 10	Comportamento geral dos critérios da manifestação cognitiva	71
Tabela 11	Comportamento geral dos critérios de utilidade	72
Tabela 12	Comportamento geral dos critérios da manifestação motivacional	73
Tabela 13	Influência dos critérios de topicalidade intelectual nos espaços de relevância	75
Tabela 14	Influência dos critérios de estrutura nos espaços de relevância	79
Tabela 15	Influência dos critérios cognitivos nos espaços de relevância	81
Tabela 16	Influência dos critérios de utilidade nos espaços de relevância	84
Tabela 17	Influência dos critérios motivacionais nos espaços de relevância	87
Tabela 18	Percepção global da relevância de acordo com a manifestação	90
Tabela 19	Dados estatísticos (<i>ex ante</i>) da Tabela 13	111
Tabela 20	Dados estatísticos (<i>ex post</i>) da Tabela 13	111
Tabela 21	Dados estatísticos (<i>ex ante</i>) da Tabela 14	112
Tabela 22	Dados estatísticos (<i>ex post</i>) da Tabela 14	112
Tabela 23	Dados estatísticos (<i>ex ante</i>) da Tabela 15	113
Tabela 24	Dados estatísticos (<i>ex post</i>) da Tabela 15	113
Tabela 25	Dados estatísticos (<i>ex ante</i>) da Tabela 16	114
Tabela 26	Dados estatísticos (<i>ex post</i>) da Tabela 16	114
Tabela 27	Dados estatísticos (<i>ex ante</i>) da Tabela 17	115
Tabela 28	Dados estatísticos (<i>ex post</i>) da Tabela 17	115

1 INTRODUÇÃO

A partir dos anos cinquenta o problema de armazenamento e recuperação de informação tem atraído a atenção de pesquisadores em múltiplas áreas do conhecimento. Com o crescimento da quantidade de informação disponível e, em particular, o advento da *internet* e a evolução das modernas tecnologias de informação, encontrar informação útil tornou-se uma das questões de investigação mais importantes nas áreas que lidam com informação (SARACEVIC, SPINK e WU, 1997). Como conseqüência, Sistemas de Recuperação de Informação (SRIs) foram e estão sendo desenvolvidos com o intuito de prestar suporte à busca de objetos informacionais (OIs).

Conforme Yao (2002), os SRIs são uma evolução dos sistemas de recuperação de dados (SRDs). A diferença entre ambos pode ser identificada na forma em que os objetos informacionais são obtidos e como a necessidade de informação do usuário é atendida. Os SRDs lidam com problemas estruturados e bem definidos como, por exemplo, os bancos de dados estruturados, onde as consultas e os métodos de associação podem ser precisamente definidos e interpretados. Os SRIs, por outro lado, originaram-se como resposta às questões envolvendo problemas informacionais dos usuários não bem definidos, onde a incerteza caracteriza o processo de recuperação de informação (RI). Essencialmente, compreendem um conjunto de procedimentos computacionais gradativamente inseridos aos clássicos SRDs.

Os SRIs têm como aspecto central a recuperação de objetos informacionais relevantes a partir de consultas efetuadas pelos usuários. As consultas são representações formalizadas das necessidades de informação dos usuários e os objetos informacionais podem ser documentos textuais ou imagens. Os critérios que os usuários utilizam para avaliar estes objetos baseiam-se no conceito de relevância, e caracterizam a importância que os objetos informacionais possam ter para o usuário (PARK, 1994).

Por conseguinte, a relevância é um tópico extremamente importante em recuperação de informação (RI), pois refere-se ao julgamento da informação recuperada em relação à necessidade de informação do usuário, ou seja, é uma medida de desempenho no processo de recuperação de informação. Esta medida é aplicada em duas situações básicas: aquela em que a relevância está relacionada a características do sistema e aquela em que a relevância está relacionada a características do usuário. Estas duas instâncias do termo têm sido designadas

como ‘relevância orientada ao sistema’ e ‘relevância orientada ao usuário’ (SCHAMBER, EISENBERG e NILAN, 1990).

A relevância orientada ao sistema refere-se a propriedades e mecanismos internos do sistema e caracteriza o resultado da correspondência entre os termos utilizados na consulta e os termos indexados e armazenados pelo sistema. A relevância orientada ao usuário inclui aspectos cognitivos, situacionais e psicológicos do usuário e refere-se aos contextos subjetivos do mesmo que são empregados para julgar os objetos informacionais (SARACEVIC, 1996a).

Estas distinções de relevância têm impacto substancial no projeto e avaliação de sistemas de recuperação de informação, uma vez que a grande maioria destes baseia-se quase que unicamente na relevância orientada ao sistema, em detrimento dos aspectos relacionados aos usuários. Conforme Draper e Dunlop (2002) e Barry (1994), a consequência desta forma como vêm sendo definidos os projetos de SRI é a evidente insatisfação dos usuários acompanhado de um relativo baixo desempenho atribuído a estes sistemas.

Neste contexto, conforme salientado por Spink e Saracevic (1998), torna-se clara a necessidade de estudos para uma consideração mais abrangente dos requisitos dos SRIs, de forma a que estes possam ser considerados como sistemas mais amplos, nos quais os usuários tenham um papel de destaque, e onde seu comportamento e atitudes possam ser levados em conta explicitamente.

Motivado por esta lacuna, este trabalho tem como objetivos estudar o comportamento dos usuários durante o processo de recuperação de informação, analisar os julgamentos de relevância efetuados pelos mesmos visando a adição destes requisitos à concepção de sistemas de recuperação de informação. Estes objetivos têm como intuito maior elucidar o processo de definição dos fatores de relevância pelos usuários de um SRI, de maneira que os desenvolvedores do sistema possam incluir no seu escopo este importante aspecto. Mais especificamente, será estudado o comportamento de usuários pertencentes ao meio acadêmico de estudos de pós-graduação.

O trabalho está estruturado em 7 capítulos, dentre os quais os capítulos 2, 3, 4, fazem parte do referencial teórico adotado para o desenvolvimento da pesquisa. O capítulo 5 descreve o método de pesquisa utilizado e no capítulo 6 são apresentados os resultados obtidos. No capítulo 7 são discutidas as considerações finais sobre esta pesquisa.

2 CONTEXTUALIZAÇÃO DA PESQUISA

As primeiras indagações sobre recuperação de informação foram imaginadas por Calvin Mooers¹ (*apud* Saracevic, 1995) na década de 50 e referiram-se às seguintes questões:

- Como representar e organizar a informação intelectualmente?
- Como especificar uma busca de forma inteligente?
- Quais sistemas e técnicas devem ser usados para dar suporte às atividades acima?

Desde esta época até atualmente, a recuperação de informação não parou de se desenvolver, tanto na teoria como na prática (SARACEVIC, SPINK e WU, 1997). Quanto à última, suas diretrizes abrangem tanto o projeto e avaliação de sistemas de recuperação de informação como o desenvolvimento de recursos computacionais (*hardware e software*) que auxiliam a confecção destes sistemas (SALTON, 1992; BORLUND, 2000). Quanto ao aspecto teórico, os esforços têm se direcionado para o usuário, tanto a partir da modelagem do processo de RI (BELKIN, 1996; SARACEVIC, 1997; INGWERSEN, 1992) como a partir dos estudos do processo de busca de informação (SPINK, 2002) e das estruturas de interação sistema-usuário (INGWERSEN, 1996).

Boa parte desta evolução deu-se no lado prático, sendo que as pesquisas voltadas para a avaliação dos sistemas de RI focaram-se basicamente nos experimentos de laboratório. Apesar deste viés ter sido útil para a evolução do aspecto computacional dos SRIs, ele, por outro lado, acentuou o problema relacionado a um dos aspectos centrais da recuperação de informação: o julgamento de relevância. Na verdade, a medida de relevância ficou reduzida ao entendimento da utilidade entre um documento e uma consulta. De acordo com Park (1994), a relevância era baseada em suposições fixas sobre o relacionamento entre o assunto de um documento e a consulta, o que ficou cunhado como ‘topicalidade’, ‘relevância tópica’ ou ‘relevância sistêmica’.

¹ MOOERS, C. Zatocoding Applied to Mechanical Organization of Knowledge. **American Documentation**, 12:131-137, 1951.

Devido a esta conjuntura, alguns trabalhos foram direcionados para ampliar o foco sobre o entendimento das características concernentes ao processo de RI. Assim, modelos foram desenvolvidos com o intuito de compreender as atividades que os usuários adotam durante o processo de RI, assim como de melhor compreender os aspectos computacionais, cognitivos, situacionais e psicológicos que estão envolvidos neste procedimento^{II}.

Decorrente do progresso destes modelos, o entendimento de relevância^{III} passou por uma revolução que levou à aceitação do fato de que uma solicitação estática não pode ser equiparada à necessidade de informação. Gradativamente passou a haver um reconhecimento sobre o caráter multidimensional e dinâmico do conceito de relevância (BORLUND, 2000) e isto ficou claro na visão de relevância articulada por Schamber, Eisenberg e Nilan (1990), onde:

1. Relevância é um conceito cognitivo multidimensional cujo significado é dependente da percepção de informação do usuário e da sua própria situação na busca por esta informação.
2. Relevância é um conceito dinâmico e que depende da qualidade do julgamento dos usuários. Este julgamento é feito a partir da observação do relacionamento entre a informação e a necessidade de informação num dado momento.
3. Relevância é um conceito complexo, porém sistemático e mensurável se abordado conceitualmente e operacionalmente pela perspectiva do usuário.

Este reconhecimento gerou diversas pesquisas na área de RI sobre a relevância, tendo como temas centrais:

- Tipos de categorias e dimensões de relevância, desenvolvidos por Saracevic (1975), Belkin (1996), Mizzaro (1998) e Cosijn e Ingwersen (2000).
- Medidas de relevância, apresentadas por Eisenberg (1998), Bruce (1994) e Tang, Vevea e Shaw (1999).
- Critérios de relevância, apresentados por Park (1993), Barry (1994), Schamber (1994) e Barry e Schamber (1998).
- Natureza da relevância, desenvolvido por Spink *et al.* (1998); Spink e Greisdorf (2001), Maglaughling e Sonnewald (2002) e Greisdorf (2002).

^{II} Estas questões são abordadas no capítulo 3 desta dissertação.

^{III} Abordado no capítulo 4 desta dissertação.

A partir destas pesquisas, houve uma remodelagem completa nas atividades de avaliação de SRI. É dentro desta problemática, e a partir da necessidade (claramente exposta pelas fontes bibliográficas) de um maior aprofundamento sobre os aspectos concernentes aos usuários, que este trabalho visa avaliar a característica central de RI: o julgamento de relevância.

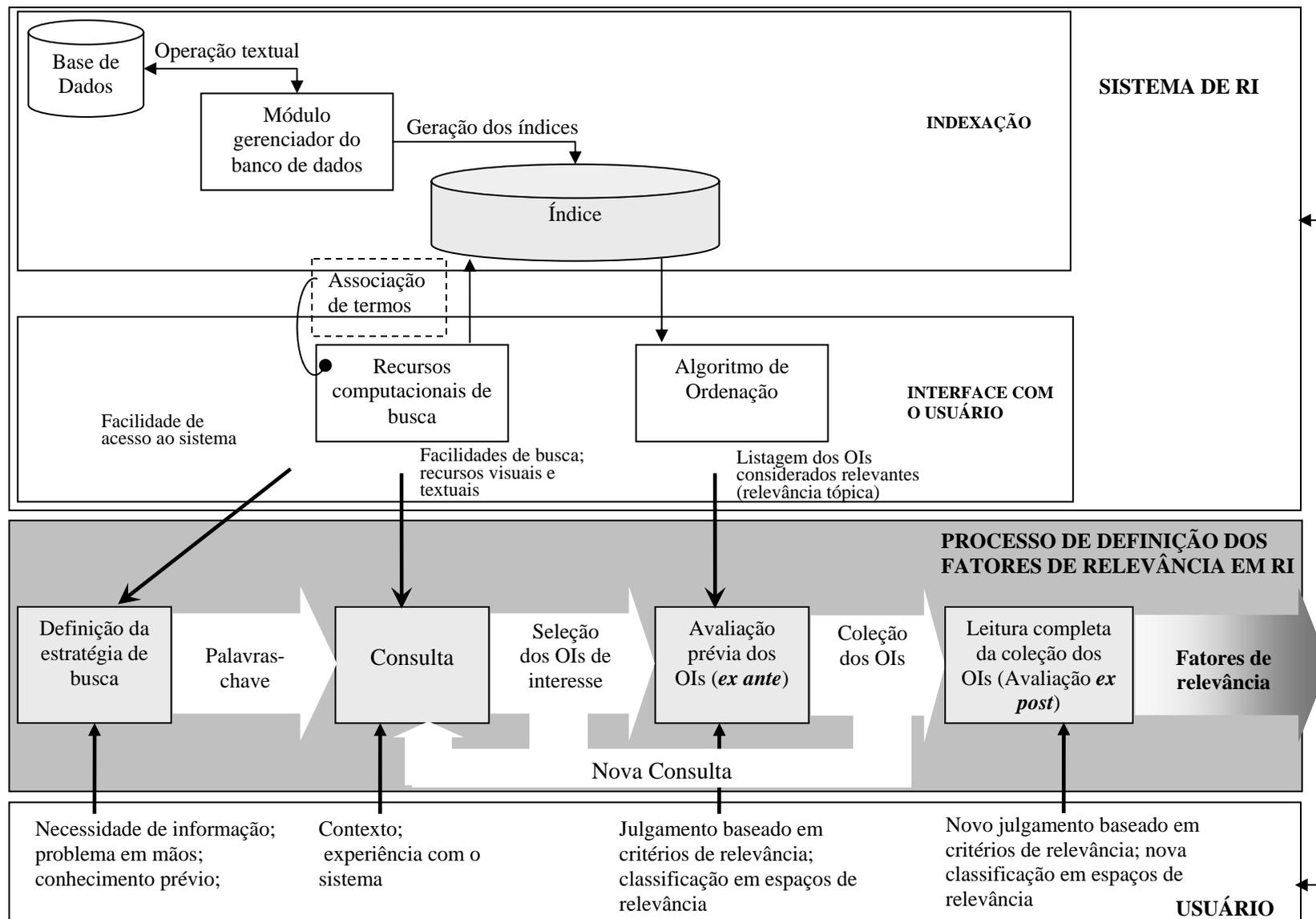


Figura 1 - Contextualização da dissertação

2.1 O PROCESSO PARA A DEFINIÇÃO DOS FATORES DE RELEVÂNCIA

Para o estudo do julgamento de relevância pelo usuário, procurou-se elucidar, com base nos estudos precedentes, o processo para a definição (pelo usuário) dos fatores de relevância na recuperação de informação, assim como os elementos que com ele interagem e que favorecem a obtenção dos resultados (Figura 1). Conforme a Figura 1, existem dois elementos principais envolvidos no processo: o sistema de RI e o usuário, sendo que o processo é fruto da interação entre ambos. Como dito anteriormente, o foco desta pesquisa está no usuário e, em especial, no processo utilizado por este no julgamento de relevância dos OIs.

A seguir uma explicação introdutória será apresentada sobre os sistemas de RI, assim como sobre a interação do processo com o usuário.

2.1.1 O Sistema de RI

Conforme Kowalski (1997), um Sistema de Recuperação de Informação é um tipo de sistema que oferece manutenção, armazenamento e recuperação da informação. O objetivo central deste tipo de sistema é prover objetos informacionais relevantes para os usuários (PARK, 1994).

Na Figura 1, o sistema de RI é dividido em dois blocos: a indexação e a interface com o usuário. Resumidamente, o processo de indexação apresenta os seguintes passos. Primeiro, é necessário a definição da *base de dados*. Isto é feito por especialistas em banco de dados, que especificam os tipos de objetos a serem utilizados e mantidos na base. Feito isto, os objetos passam por uma operação textual que tem como finalidade adaptá-los a um tratamento lógico, gerando assim metadados. Estes correspondem às informações sobre a organização dos dados dos OIs, sobre quais domínios estes dados pertencem e sobre o relacionamento entre eles. Após esta tarefa, o *módulo gerenciador do banco de dados* especifica os esquemas dos metadados (relações, campos e atributos de cada relação, etc.) e constrói um *índice* dos objetos informacionais. Conforme Baeza-Yates e Ribeiro-Neto (1999), o índice é uma estrutura de dados importantíssima para o sistema, porque permite uma busca rápida dentro de bases volumosas.

Quanto à interface do sistema com o usuário, pode ser entendida como um programa que interage com o usuário e controla a base de recursos informacionais do SRI. Tem como objetivo principal capacitar o usuário a desenvolver sua atividade de forma eficaz e autônoma (sem a necessidade de intervenção de especialistas) (BRAJNIK, MIZZARO e TASSO, 1996). Para o usuário, a interface oferece recursos tais como menus, janelas, ícones, gráficos, caixas de diálogos e recursos de busca como, por exemplo, procura por tipo específico de OI e por atributos informacionais (título, *abstract*), dentre outros.

Como a interface é um programa, ela também é responsável por uma parte do processamento da consulta, ordenação e visualização dos OIs recuperados da base indexada. Certamente, a interface é de extrema importância para o processo de RI, pois além de dar suporte ao usuário, livra-o de características computacionais inerentes ao sistema.

2.1.2 O Usuário e o Processo

O processo para a definição (pelo usuário) dos fatores de relevância (Figura 1) tem início quando o usuário possui uma necessidade de informação e procura um SRI para dar suporte à busca de informação (relevante) e que o auxilie na resolução do seu problema informacional. Inicialmente, é necessário que o usuário defina a *estratégia de busca de informação*, ou seja, como será feita a representação implícita do problema em mãos de forma compatível com os recursos de processamento que o sistema oferece. Esta etapa é restringida pelo conhecimento prévio que o usuário possui sobre o assunto e pelas possibilidades que a interface do sistema apresenta para o acesso à informação. Definida a estratégia de busca, obtém-se as palavras-chave a serem utilizadas na consulta.

Uma *consulta* constitui a representação formal da estratégia de busca. É influenciada pelo contexto que o usuário se insere e pela experiência que ele possui em interagir com o sistema de RI. A interface do sistema oferece facilidades e restrições de busca, como por exemplo, conectivos *and* e *or*, pesquisa por campos do OI, asteriscos para substituir letras em palavras, gráficos de estatística dos documentos, refino de buscas dentro da primeira coleção de OIs recuperados, entre outros. Durante a consulta, o usuário seleciona os objetos informacionais recuperados que aparentam ser do seu interesse.

A etapa de *avaliação prévia dos OIs (ex ante)* compreende a análise superficial pelos usuários do conteúdo dos OIs recuperados e que aparentam ser do seu interesse. Caso o usuário não se dê por satisfeito, novas consultas podem ser feitas. Nesta etapa o usuário,

baseando-se no quão relevante cada objeto foi para sua necessidade de informação, classifico em uma escala de relevância.

A última etapa, diz respeito a um novo julgamento de cada um dos OIs classificados de acordo com sua ordem de relevância. Este novo julgamento (*avaliação dos OIs ex post*), é baseado não mais numa leitura superficial, e sim na leitura completa de cada documento. Isto possibilita que seja verificada efetivamente a mudança de percepção do usuário em relação ao OI anterior e, principalmente, a eficácia do sistema em retornar objetos relevantes (sua função essencial). Como resultado desta etapa, é possível determinar com maior exatidão os fatores de relevância que influenciam o usuário no seu julgamento dos objetos informacionais, o que é de extrema importância para os projetos de sistemas de RI. Neste trabalho, entende-se fatores de relevância como a correlação e influência dos critérios com as regiões ou espaços de relevância.

Uma vez evidenciado o processo que constitui o objeto de estudo desta pesquisa, procuraremos identificar a situação problemática que ela visa abordar e, por conseguinte, os objetivos a serem atingidos.

2.2 SITUAÇÃO PROBLEMÁTICA

Conforme Saracevic (1995), as características dos SRIs que favorecem sua avaliação podem ser agrupadas em:

1. **Engenharia:** Nesta classe, as questões de avaliação são dirigidas ao desempenho de *hardware* e de *software*, como segurança, erros, tolerância a falhas, rapidez, integridade, sustentabilidade e flexibilidade. Para esta classe, a efetividade computacional e a eficiência dos métodos e algoritmos de recuperação são investigadas.
2. **Entradas:** Aqui, questões referentes às entradas e conteúdos do sistema são investigadas.
3. **Processamento:** Nesta classe, indagações sobre os meios de processamento das entradas caracterizam o processo de avaliação. Isto inclui avaliação do desempenho dos algoritmos e das técnicas computacionais empregadas na construção do sistema.

4. **Saídas:** Questões sobre a interação com o sistema e as saídas obtidas são levantadas. Isto inclui avaliação do processo de busca, interação, *feedback*, saídas, etc.
5. **Uso e usuário:** Questões relacionadas ao tratamento dos problemas em mãos do usuário e respectivas tarefas são salientadas.
6. **Social:** Esta classe refere-se a assuntos relacionados ao impacto no ambiente, onde o processo ocorre como, por exemplo, produtividade, tomada de decisão, etc.

Como dito anteriormente, um dos problemas atuais em recuperação de informação é a concentração de estudos direcionados aos aspectos do sistema e, em especial, em níveis de engenharia e processamento, sendo dada pouca atenção aos níveis que envolvem o usuário. Devido a isto, embora os programas tornem-se cada vez mais sofisticados, os sistemas de RI caminham em uma trilha inadequada (SARACEVIC, 1995).

Evidencia-se isto à luz do tratamento estático dado à relevância, onde a tal classificação é feita de forma dicotômica - relevante ou não relevante – e constitui um aspecto prejudicial aos projetos de SRI e tem sido empregada com muita frequência para a definição do seu escopo. Conforme Spink, Greisdorf e Bateman (1998), mesmo as técnicas de associação mais potentes e atuais que provêm uma ordenação de relevância calculada de forma probabilística ainda assumem esta decisão dicotômica como um ponto de corte.

Estudos que incluam a consideração de espaços (regiões) parciais de relevância permitiriam uma maior clareza e coerência na especificação e desenvolvimento de projetos de sistemas de recuperação de informação, tornando-os mais efetivos (SPINK e GREISDORF, 2001). Isto porque a natureza das técnicas de ordenação e ponderação dos OIs durante o processo de recuperação de informação pode ser significativamente aprimorada com o entendimento das decisões referentes às regiões parciais de relevância.

Outro problema atual em RI refere-se ao fato de que as pesquisas estão progredindo em duas áreas desconectadas. Por um lado, tem-se focado o nível e os critérios de relevância e, por outro, suas regiões. Conforme Borlund (2003), uma possibilidade interessante de unir as pesquisas seria verificar a existência de correlação entre critérios e regiões estudados por alguns pesquisadores (denominado anteriormente como fatores de relevância) e os tipos de relevância descritos por Saracevic (1996b). Desta forma obter-se-ia um avanço significativo no entendimento do comportamento do usuário durante o processo de julgamento de

relevância e, conseqüentemente, um aprimoramento nos procedimentos de RI e desenvolvimento de sistemas de recuperação de informação.

Em suma, uma compreensão completa dos requisitos de um sistema é fundamental para o desenvolvimento bem sucedido do mesmo. A definição dos requisitos oferece ao desenvolvedor e ao cliente os critérios para avaliar a qualidade do sistema, durante e após sua construção (PRESSMAN, 1995). A consideração explícita dos requisitos de relevância possibilita que o engenheiro de sistemas especifique melhor a função e o desempenho do sistema, indique com maior visão de conjunto a interface do *software* com outros elementos do sistema e estabeleça com maior riqueza quais são as restrições de projeto que o sistema de RI deve enfrentar.

Tendo em conta os fatos acima elucidados, este trabalho focaliza o usuário e os aspectos (critérios e regiões) de relevância de que se serve para avaliar os objetos informacionais recuperados pelos SRIs, e considera no seu desenvolvimento os componentes do processo apresentado na Figura 1. A seguir apresentam-se os objetivos cujo alcance é esperado nesta pesquisa.

2.3 OBJETIVOS

Com base no processo da Figura 1 e nos problemas expostos anteriormente, os objetivos desta pesquisa podem ser apresentados da seguinte maneira:

OBJETIVO GERAL

Caracterizar os julgamentos dos usuários em espaços de relevância, definidos através da consideração dos critérios, regiões e manifestações de relevância de que se servem para avaliar os objetos informacionais recuperados em um SRI. Espera-se que a caracterização dos julgamentos contribua para a especificação de requisitos para projetos, avaliação e desenvolvimento de Sistemas de Recuperação de Informação.

OBJETIVOS ESPECÍFICOS

I. Analisar a formulação das estratégias de busca durante o processo de recuperação de informação.

- II. Analisar o desenvolvimento das consultas durante o processo de RI.
- III. Levantar novos critérios utilizados pelos usuários.

3 RECUPERAÇÃO DE INFORMAÇÃO

Recuperação de informação é uma atividade complexa que envolve diversas ações, como a organização de objetos informacionais (documentos textuais, arquivos multimídia), a representação intelectual dos OIs derivados diretamente por pessoas ou indiretamente por algoritmos, a busca e recuperação intelectual feita pelos usuários e sistemas e técnicas computacionais para suportar estas ações (SPINK e SARACEVIC, 1998). Mais especificamente, esta complexidade acentuou-se a partir dos anos setenta, quando a simbiose entre a computação e a telecomunicação tornou os SRI interativos. (SARACEVIC, 1997).

Conforme Baeza-Yates e Ribeiro-Neto (1999), a área de recuperação de informação tem tido um crescimento que vai muito além dos seus objetivos primários, que eram a indexação de textos e a busca por documentos úteis. Atualmente, pesquisas em RI incluem modelagem, classificação de documentos e categorização, arquitetura de sistemas, interfaces com usuário, visualização de dados, filtragem e linguagem. Boa parte destes temas de pesquisa foi aprofundado com o aparecimento da Internet.

No entanto, o advento desta tecnologia produziu alguns efeitos colaterais. Devido à imensa quantidade de informação, o aspecto primordial da recuperação de informação, que é a descoberta de informação relevante aos requisitos de informação desejados pelos usuários, tornou-se uma tarefa difícil, frequentemente produzindo resultados com baixa qualidade (BAEZA-YATES e RIBEIRO-NETO, 1999). Conforme Barry (1994), este problema acontece principalmente devido à disparidade existente entre os projetos de sistemas de recuperação de informação e as medidas de avaliação de relevância aplicada pelos usuários.

Na verdade, as características provenientes da interação entre o usuário e o sistema durante a recuperação de informação não fazem parte do esforço primário em grandes projetos de SRI, sejam eles envolvendo desenvolvimento e/ ou avaliação (SARACEVIC, 1997).

Em suma, o que deve estar em jogo são os aspectos da interação entre o sistema e o usuário e, em especial, os fatores de relevância. Para isto acontecer, as características que envolvem o processo de recuperação de informação devem estar bem especificadas. A seção seguinte discute estas características e os modelos principais que abordam o processo de RI.

3.1 PROCESSO E MODELOS DE RECUPERAÇÃO DE INFORMAÇÃO

O processo de recuperação de informação advém da necessidade do usuário obter resposta para a resolução de algum problema informacional. Conforme Bruce (1994), este processo possui três fases:

- Reconhecimento do problema: esta fase compreende o problema em si, o conhecimento interno do problema e a intenção informacional. Nesta etapa, as estruturas pré-concebidas de conhecimento do usuário são consideradas como suficientes para a noção do que é relevante ou não para o problema em questão (SARACEVIC et al., 1988).
- Interação com o sistema: táticas de buscas são estabelecidas para enfrentar o problema informacional e são alteradas ou validadas de acordo com as respostas retornadas pelo sistema de RI. Esta interação é considerada como um processo de aprendizagem que envolve o desenvolvimento e alteração do estado do problema.
- Interação com o documento: refere-se ao ponto de acesso e interação com o(s) documento(s) recuperados.

Inicialmente, para visualizar as duas primeiras fases, apresenta-se abaixo a Figura 2 que ilustra as características sob o ponto de vista do usuário. Em primeira instância, o estado do problema é caracterizado pela percepção de uma situação problemática e a necessidade de informação (NI) para resolvê-la.

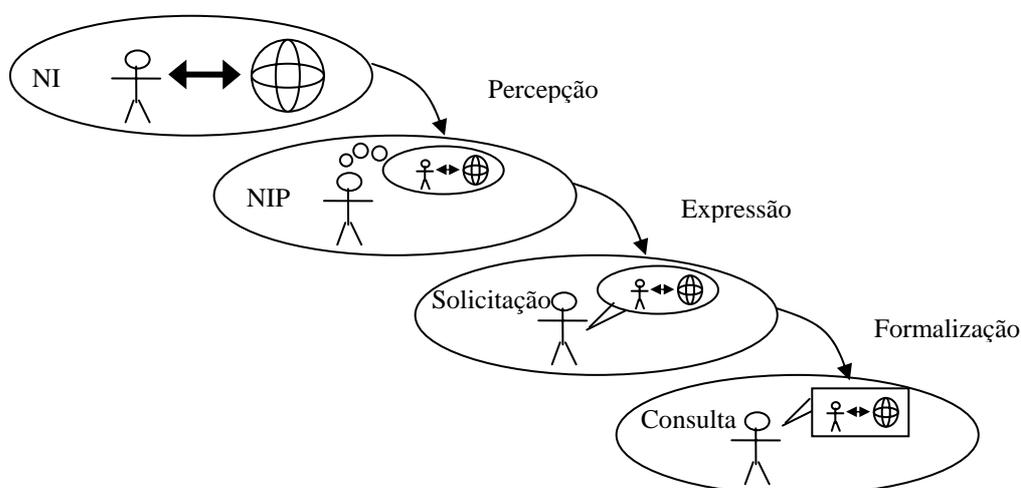


Figura 2 - Processo sob o ponto de vista do usuário
Fonte: MIZZARO, 1998.

Feito isto, o usuário constrói uma representação implícita da situação problemática, que corresponde na Figura à necessidade de informação percebida (NIP). A partir disto, o usuário expressa sua NIP em uma solicitação (normalmente feita em linguagem natural) e, por fim, formaliza a solicitação em uma consulta para o sistema de recuperação de informação (MIZZARO, 1998). Já, sob o ponto de vista do sistema, suas características podem ser resumidas na Figura 3.

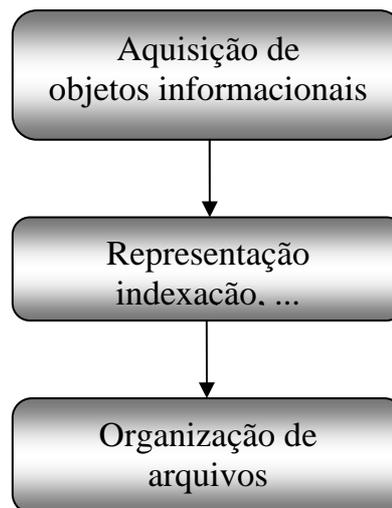


Figura 3 - Processo sob o ponto de vista do sistema
Fonte: SARACEVIC, 1996a.

Inicialmente, os OIs (textos, imagens, dados) existem em uma base de dados que é freqüentemente atualizada. Nesta mesma base, estes objetos passam por uma operação textual que os transforma em uma representação lógica, e que servirá para a criação dos índices e a organização dos OIs na base.

Unindo o processo sob o ponto de vista do usuário ao da Figura 3, obtém-se o modelo tradicional de RI. Neste modelo, o ponto de contato entre o usuário e o sistema dá-se no momento em que o usuário solicita o processamento de sua consulta. Esta conjunção pode ser visualizada na Figura 4.

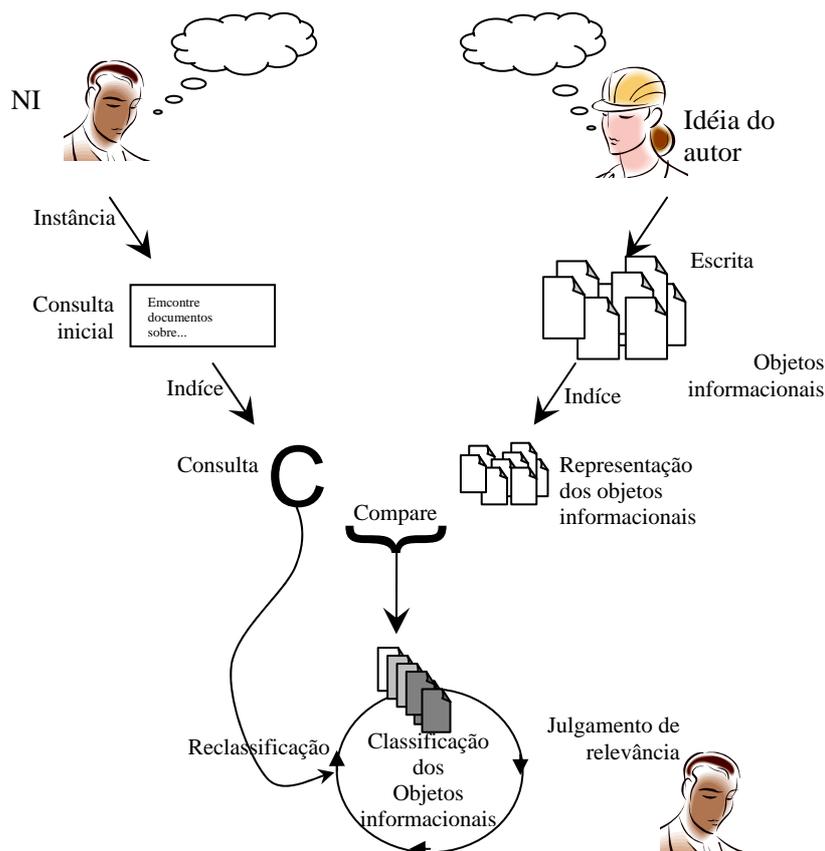


Figura 4 - Modelo Tradicional de RI
 Fonte: DUNLOP, JOHNSON e REID, 1998.

Neste modelo são considerados tanto os aspectos do sistema de recuperação como a intervenção feita pelo usuário sobre os objetos informacionais recuperados pelo sistema. Inicialmente neste processo, o autor confecciona um objeto informacional e cada objeto é armazenado no sistema de RI. Normalmente, a representação destes OIs é feita mediante a utilização de palavras-chave.

O processo de RI inicia-se quando o usuário instancia sua necessidade de informação em uma consulta (estado do problema e interação com o sistema). Os termos desta consulta são comparados pelo sistema com as representações dos objetos informacionais na base interna do sistema. O processo de comparação resulta em uma classificação dos objetos de acordo com uma estimativa feita pelo sistema de RI de quão próximo cada objeto corresponde à consulta efetuada. Decorrente desta classificação, os usuários julgam os objetos informacionais quanto à sua relevância. E, por fim, a reclassificação preocupa-se com a lista de documentos recuperados e serve para que o usuário relacione o que ele considera como relevante ou para que ele modifique sua consulta em face da sua necessidade de informação.

Este modelo tem como vantagem conseguir o isolamento das variáveis envolvidas no lado do sistema e permitir uma concentração uniforme de suas aplicações, avaliações e análises (SARACEVIC, 1997). Por consequência desta atividade, houve um avanço considerável das técnicas computacionais como, por exemplo, estudos das representações textuais que envolvem classificação, indexação, processamento de linguagem natural e modelos para ordenação de documentos, sendo que muito destas pesquisas estão sendo utilizadas e/ ou integradas a partir de técnicas de inteligência artificial e de automação industrial (INGWERSEN, 1992).

No entanto, apesar do modelo tradicional ser de grande utilidade para projetos de SRI, existem algumas críticas contundentes sobre a sua efetividade, como por exemplo:

- Há uma grande ênfase sobre o processo de representação textual e, relativamente, pouco esforço sendo utilizado no desenvolvimento de esquemas representativos das necessidades de informação ou problemas. Mais especificamente, há pouca preocupação com o processo de julgamento de relevância que é tipicamente aplicado pelo usuário (BELKIN, 1993).
- Como consequência direta do problema acima, este processo de julgamento de relevância é usado pelo sistema para unicamente modificar a consulta ou ocasionalmente modificar as representações textuais. Ou seja, a relevância é considerada apenas sob o ponto de vista do sistema e, portanto, entendida como estática (associação entre os termos da consulta e o assunto de um OI) (KEKÄLÄINEN e JÄRVELIN, 2002).
- Unindo os dois problemas acima, obtém-se o mais agravante. Neste modelo, quase nenhum aspecto de interação é incorporado (SARACEVIC, 1997). Ou melhor, o envolvimento do usuário com o sistema é considerado mínimo e a interação como o processo de julgamento é vista, simplesmente, como suporte para o processo de associação e comparação (BELKIN, 1993).

Em suma, boa parte da pesquisa dedicada ao modelo tradicional limita-se ao desenvolvimento de algoritmos e à comparação do desempenho entre eles, buscando caminhos para otimizá-los. No aspecto teórico, os objetivos incluem a análise básica de problemas de RI, tais como a representação de documentos e consultas e os estudos de métodos de associação (KEKÄLÄINEN e JÄRVELIN, 2002).

Em contraste com os aspectos anteriormente elucidados, Ingwersen (1992, 1996) propôs um modelo que se concentra na identificação dos processos cognitivos que podem ocorrer em todos elementos de processamento de informação. A Figura 5 apresenta estes processos.

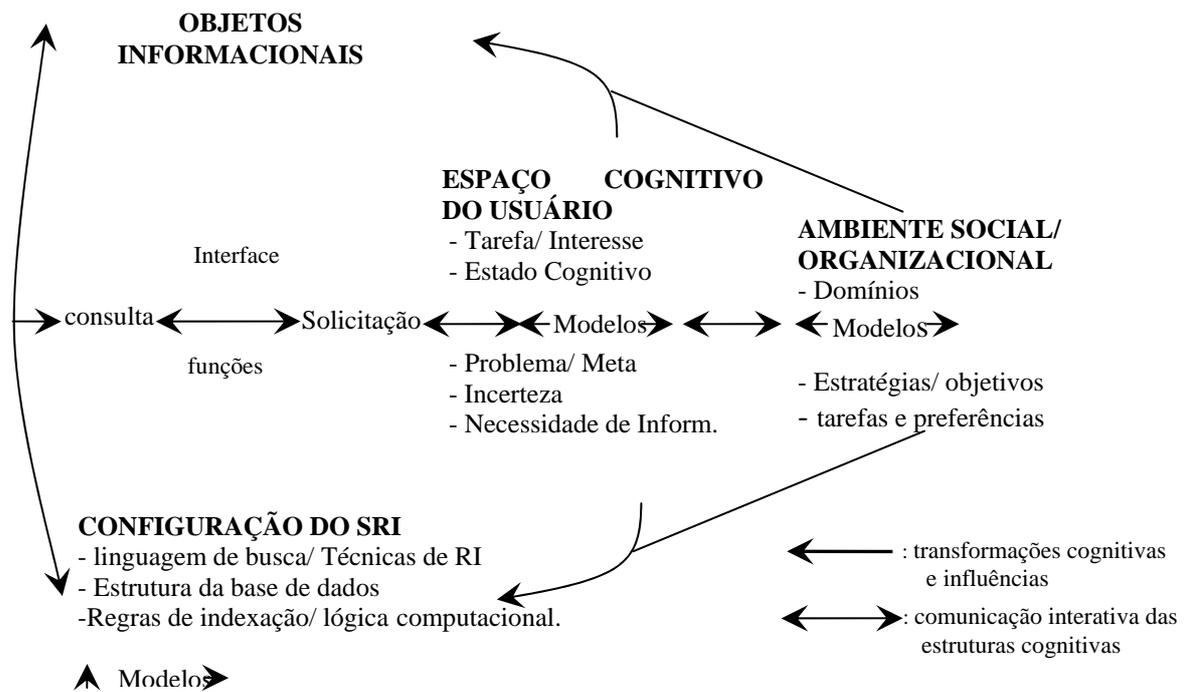


Figura 5 - Modelo cognitivo de interação em RI
Fonte: INGWERSEN, 1996.

Na Figura acima, fica evidente que o modelo é fundamentado em quatro âncoras. A primeira delas é a configuração do SRI, que consiste em estruturas cognitivas dos projetistas do sistema. Isto é, corresponde à visão da concepção do projetista de como são processados os OIs dentro do sistema (INGWERSEN, 1992). A segunda âncora diz respeito aos objetos informacionais. Estes OIs podem ser representações do conhecimento, tais como textos completos, figuras, entre outros. Os objetos relacionam-se com as configurações do sistema a partir de regras de relevância ou algoritmos. Conforme Ingwersen (1996), este relacionamento pode alcançar um nível cognitivo apenas no caso de uma intervenção por um indexador humano.

Obviamente, os OIs são acessados por consultas que são solicitadas de acordo com as funções disponíveis na interface do sistema. A estruturação da solicitação é fortemente influenciada pela terceira âncora, o espaço cognitivo do usuário. Neste, estão envolvidas as estruturas de conhecimento do usuário, metas e desejo de informação, que instiga a atividade

de RI. Certamente, a estrutura cognitiva do usuário influencia e é influenciada pelo ambiente social e organizacional (quarta âncora) em que ele está inserido (INGWERSEN, 1996). Isto porque preferências, estruturas cognitivas coletivas e convenções sociais aderem aos domínios (setores industriais, firmas, organizações, comunidades de prática).

Sucintamente, este modelo pode ser resumido nos seguintes pontos principais (SARACEVIC, 1996a):

- A interação durante a recuperação de informação compreende uma série de processos cognitivos, ocorrendo potencialmente em todos os elementos de processamento de informação.
- O espaço cognitivo do usuário compreende uma série de elementos causais estruturados, onde contextos cognitivos e situacionais predominam.
- A interação ocorre em diferentes níveis e, conseqüentemente, em diversas instâncias.
- O processo é altamente dinâmico. Uma múltipla representação é aplicada tanto para o espaço cognitivo do usuário quanto para o espaço de informação do sistema.

Comparativamente, este modelo (Figura 5) é mais robusto do que o modelo tradicional (Figura 4) porque se esforça em descrever a interatividade entre o usuário e o sistema e a influência que ambos exercem entre si. No entanto, conforme Saracevic (1996a), o ponto negativo deste modelo diz respeito ao aspecto referente à sua testabilidade (principalmente, os dois últimos pontos principais citados acima) e sua aplicação em avaliação de SRI.

Também como resposta ao modelo tradicional de RI, foi criado um novo modelo onde o SRI deve ser composto pelos componentes dos usuários e pelos recursos de conhecimento destes, sendo que ambos são unidos por eventos específicos. Na Figura 6 a seguir podem ser visualizados estes eventos, o processo e os componentes deste modelo.

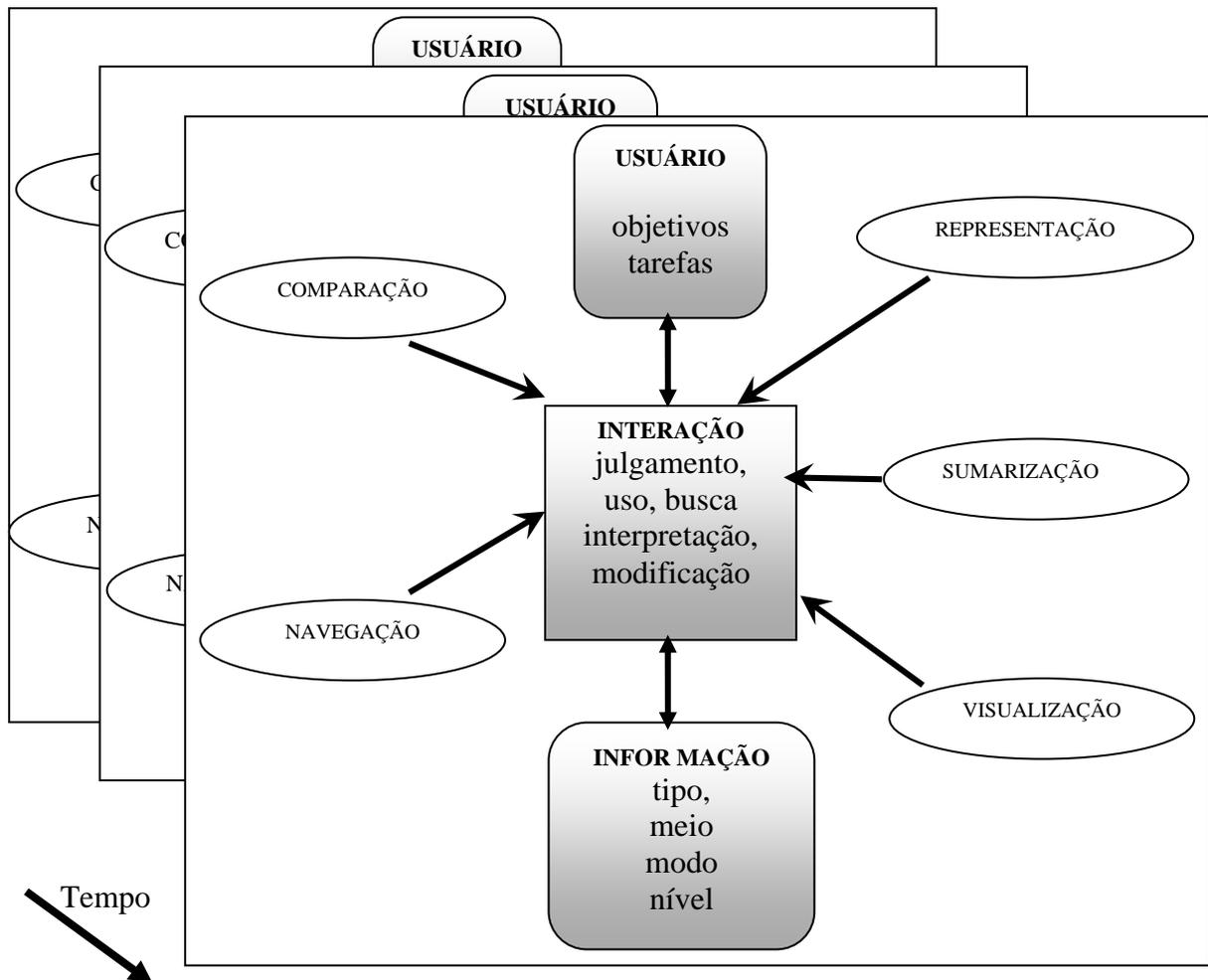


Figura 6 - Modelo de interação em RI baseado em eventos
 Fonte: BELKIN, 1996.

Conforme Belkin (1996), o modelo baseado em eventos consiste em uma série de tipos de interações que são estruturadas de acordo com os objetivos, intenções, tarefas e metas a serem atingidas pelos usuários ao longo do tempo, sendo que estes objetivos podem ser afetados pelos componentes da interação. Neste modelo os processos tradicionais de RI (representação, visualização, sumarização, navegação e comparação) servem apenas de suporte ao processo interativo do usuário com a informação (BELKIN, 1996). Esta interatividade com a informação é baseada em eventos, dos quais fazem parte a interpretação do conteúdo de cada OI, os julgamentos referentes à qualidade dos objetos informacionais e sua utilidade, entre outros.

Em suma, o que está em jogo é a interação do usuário com a informação e não com o sistema (BELKIN, 1993). Mais especificamente, o que interessa é a representação dos estados

irregulares ou anômalos de conhecimento do usuário e os aspectos cognitivos e situacionais que são a razão para a busca de informação e caracterização dos SRIs.

O aspecto importante do modelo baseado em eventos é o seu direcionamento à uma variedade de processos em RI, diferentemente do modelo tradicional que é focado de forma preponderante nos aspectos de associação, como já mostrado anteriormente. No entanto, há uma dificuldade em identificar os eventos individuais e seus efeitos nos outros.

Este foi um dos aspectos que fez Saracevic (1996a) propor um modelo baseado em camadas onde é tentado: A) reforçar e otimizar os aspectos positivos dos modelos anteriores, B) minimizar as deficiências tanto do modelo focado no sistema (Figura 4) quanto dos modelos focados no usuário (Figuras 5 e 6), e C) criar uma estrutura que considere a natureza da relevância em RI. A Figura 7 abaixo exhibe a estrutura do modelo.

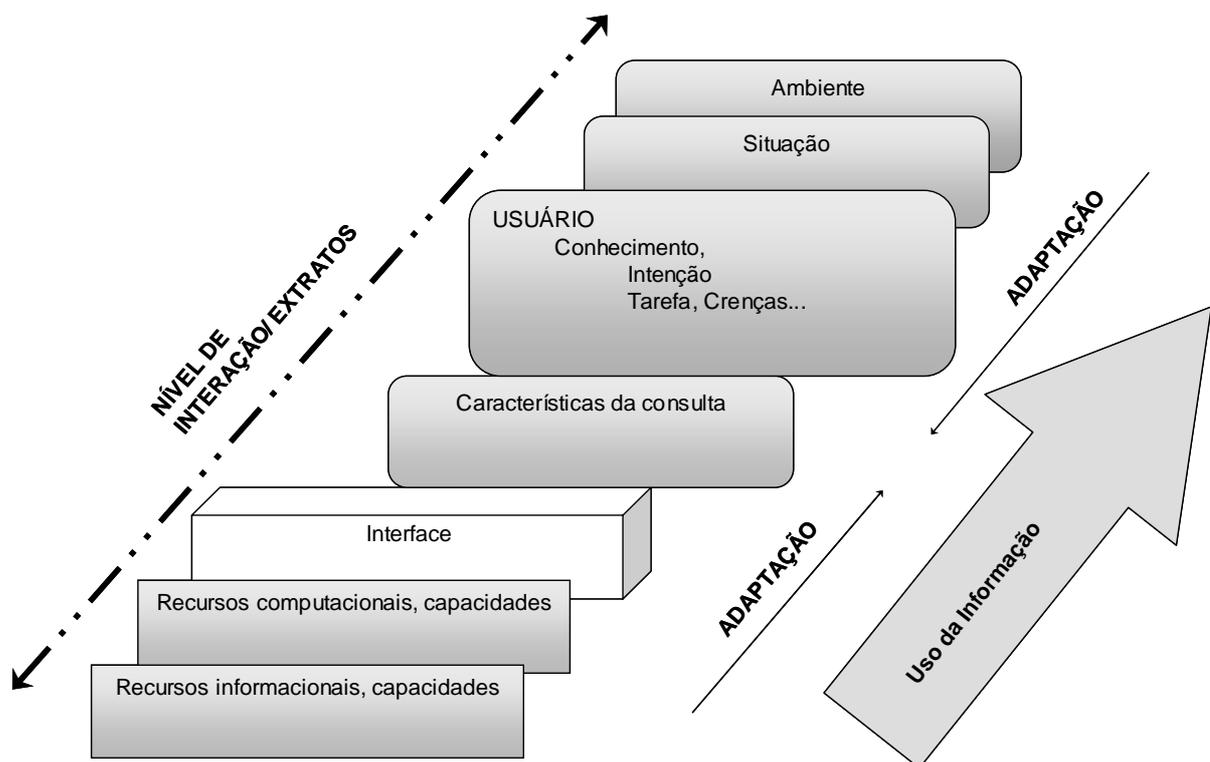


Figura 7 - Modelo de interação em RI baseado em camadas
Fonte: SARACEVIC, 1996b.

O modelo baseia-se em três atividades principais:

- Aquisição: envolve a obtenção de informação, reconhecendo que as informações podem ser de vários tipos (SARACEVIC, 1997).
- Cognição: envolve a absorção de informação e seu processamento cognitivo (SARACEVIC, 1996a).

- **Aplicação:** refere-se ao uso da informação adquirida e absorvida para a tarefa que o usuário tem em mãos, levando em conta o ambiente e a situação em que se encontra o usuário.

Cada um dos componentes acima possui elementos distintos com diferentes papéis, propósitos, processos, adaptações, entre outros. Desta forma, fica evidente na Figura 6 que o processo de RI é um diálogo entre dois participantes principais (usuário e computador). Este diálogo ou interação compreende uma seqüência de processos que ocorrem em níveis ou camadas através de uma interface. O quadro abaixo exhibe os componentes do modelo:

Quadro 1 - Componentes e níveis do modelo em camadas Fonte: SARECEVIC, 1997.

Elementos do modelo	Níveis
Humano	Cognitivo Afetivo Situacional
Computador	Engenharia Processamento Conteúdo
Interface	Superfície

De acordo com Saracevic (1997), o lado humano possui três níveis de interação:

- **Cognitivo:** neste, o usuário interage com o OIs, interpretando-os e julgando-os para assimilá-los cognitivamente. Este nível concentra processos e resultados cognitivos, como julgamento de relevância e efeitos e/ ou mudanças no estado de conhecimento do usuário.
- **Afetivo:** os usuários trazem para o processo de RI suas intenções, crenças, motivações, sentimentos e desejos.
- **Situacional:** os problemas e situações dos usuários produzem suas necessidades de informação e questionamentos durante o processo de RI. Neste nível, os objetos são analisados de acordo com sua utilidade.

Obviamente, os fatos são mais complexos do que exposto acima. Muitas vezes, os problemas que os usuários possuem são mal definidos. Isto interfere diretamente na especificação dos termos de consulta, nas táticas de busca e na seleção de objetos

informacionais. Por conseguinte, questões podem ser redefinidas e redimensionadas e consultas podem ser abandonadas e novas serem feitas, sendo que estas mudanças são materializadas no nível de superfície. Este nível pertence à interface do sistema (mostrada na Figura 7) e possui uma série de eventos, tanto do lado usuário quanto do computador.

Estes eventos envolvem as definições das expressões (comandos) de busca efetuadas durante o processo da consulta ao sistema, a aprendizagem e entendimento dos recursos informacionais disponibilizados pelo sistema (como, por exemplo, opções de navegação e visualização de resultados), a reestruturação de consultas, a utilização de algoritmos e procedimentos que processam as transações de consultas e devolvem os resultados na interface do sistema, e a inferência da relevância dos OIs (SARACEVIC, 1996a). Dando suporte aos aspectos dos usuários e da interface, estão os recursos computacionais, que possuem os seguintes níveis:

- Engenharia: abrange o *hardware* e seus atributos operacionais como capacidade, eficiência e poder de processamento.
- Processamento: concentra-se nos vários tipos de *software*, especialmente os de manipulação textual, consulta e projeto de interface e estruturação de arquivos.
- Conteúdo: preocupa-se com as representações informacionais. Ou seja, inclui coleções de meta-informação e/ ou meta-característica de cada OI.

De forma geral, este modelo é um dos mais robustos em RI, mesmo considerando-se que a decomposição e a interação entre as camadas não são tarefas fáceis de serem executadas. A preocupação em incluir os aspectos do sistema (modelo tradicional) e os aspectos do usuário (modelo cognitivo e modelo em eventos) favorece um maior entendimento global dos aspectos concernentes ao processo de RI e, em especial para este trabalho, dos fatores de relevância. Isto porque as manifestações de relevância existem em todas as camadas do modelo e são interdependentes (SARACEVIC, 1996a). Assim, fica evidente que o julgamento acerca de relevância dos objetos informacionais é um dos fatores mais importantes no processo de RI, porque relaciona-se diretamente com o propósito central da recuperação de informação, que é a obtenção de objetos informacionais que auxiliem os usuários a resolver seus problemas.

Objetivamente, tanto o processo de julgamento de relevância como o comportamento do usuário no processo de RI têm profundo impacto no projeto de sistemas de recuperação de informação. Conforme Barry (1994), a incorporação dos fatores de relevância dos usuários ao

projeto dos sistemas seria uma via possível para melhorar o desempenho destes sistemas. Desta forma, faz-se necessário um melhor entendimento do conceito de relevância neste processo e na avaliação de sistemas de RI, assunto do próximo capítulo.

4 RELEVÂNCIA

As medidas mais utilizadas como parâmetro para avaliação do desenvolvimento e do desempenho de sistemas de RI têm sido aquelas designadas como *recall* e *precision*.



Figura 8 - Efeitos de busca no espaço total da base de dados
Fonte: KOWALSKY, 1997.

Conforme Su (1994), ambas as medidas são definidas da seguinte forma (ver Figura 8):

$\textit{Recall} = \frac{\text{n}^\circ \text{ de objetos informacionais relevantes recuperados}}{\text{n}^\circ \text{ total de objetos informacionais relevantes na base de dados}}$
$\textit{Precision} = \frac{\text{n}^\circ \text{ de objetos informacionais relevantes recuperados}}{\text{n}^\circ \text{ total de objetos informacionais recuperados na base de dados}}$

onde :

- O número de objetos informacionais relevantes recuperados refere-se à quantidade de objetos informacionais relevantes que foram recuperados pelo SRI, dentre todos os objetos relevantes constantes na base.
- O número total de objetos informacionais relevantes na base de dados refere-se à quantidade máxima de objetos informacionais relevantes que a base de dados contém (relevantes recuperados + relevantes não recuperados).
- O número total de objetos informacionais recuperados na base de dados refere-se à quantidade de objetos informacionais que foram recuperados pelo sistema.

Constam, entretanto, na literatura, uma série de objeções a estas medidas. Primeiramente, *recall* e *precision* parecem ter uma relação inversa entre si. Segundo, conforme Su (1992), *precision* é relativamente simples de ser aplicada, mas *recall* é extremamente difícil de ser implementada porque é requerido o conhecimento de todos os objetos informacionais relevantes numa base de dados, ou seja, é praticamente impossível aplicá-la em sistemas que atuem em bases de dados volumosas (SALTON, 1992). Dentre estas críticas, a mais agravante é que a segunda medida (*precision*), uma importante e tradicional medida de efetividade, não é significativamente correlacionada com os julgamentos dos usuários quanto ao sucesso dos objetos informacionais recuperados no processo de RI. De acordo com Su (1994), isto sugere que esta medida não é um bom indicador de sucesso.

Na verdade, estas medidas seguem completamente o modelo tradicional de RI (ver Figura 4) e, por conseqüência, referem-se apenas à eficiência dos procedimentos computacionais do sistema; mais especificamente, aos procedimentos para comparação da consulta com a representação dos objetos informacionais e para a classificação destes objetos. Como visto nos demais modelos de RI, os usuários do sistema são peças-chave no processo de recuperação de informação, uma vez que uma das características mais importantes deste processo é o julgamento de relevância efetuado por estes usuários. A partir disto, os pesquisadores da área dedicaram especial atenção ao conceito de relevância porque este é mais completo do que os outros dois aqui citados e é um parâmetro de sucesso amplamente utilizado pelos usuários de SRI. Assim, a próxima seção visa esclarecer o que está envolvido neste conceito.

4.1 A EVOLUÇÃO DO CONCEITO DE RELEVÂNCIA

O conceito de relevância têm sido discutido desde a década de 50. Um dos primeiros conceitos de relevância discutidos na literatura foi o definido por Cuadra e Katter^{IV} (*apud* Saracevic, 1975). Conforme estes autores, a relevância era a correspondência, em um contexto, entre uma declaração de informação requisitada por uma consulta e algum documento contido na base de dados. Neste caso, o que está em jogo é a relação entre o conteúdo coberto pelo documento e o conteúdo associado à declaração de informação requisitada.

Outra variante desta definição afirma que a relevância é uma medida de utilidade existente entre um documento e uma consulta, julgada pelo requerente desta consulta (SARACEVIC, 1970). O contexto destas definições advém dos experimentos fechados em laboratório que utilizam consultas pré-definidas e conhecimento prévio dos documentos relevantes, tendo como objetivo principal a avaliação do desempenho de algoritmos e técnicas computacionais (SARACEVIC, 1995).

O exemplo clássico deste contexto e visão de relevância foi o experimento realizado por Cleverdon e Keen^V (*apud* Hersh, 1994). Neste trabalho, os autores criaram uma coleção de testes com consultas pré-fixadas, documentos e julgamentos de relevância para verificar o desempenho do sistema. Em nenhum momento houve envolvimento dos usuários com suas opiniões quanto ao desempenho do sistema de RI envolvido neste ensaio (HERSH, 1994).

Experimentos semelhantes a este reduzem e abordam a relevância sob o ponto de vista da idéia de topicalidade. Em uma medida de topicalidade, a relevância restringe-se a uma relação entre um tópico/ tema abordado por um documento recuperado e a consulta efetuada pelo usuário. Relevância vista desta forma contém uma visão extremamente mecanicista dos fenômenos envolvidos no processo de recuperação de informação, porque não levam em conta os aspectos cognitivos e os critérios empregados pelos usuários no seu julgamento. Esta percepção de relevância foi muito criticada, principalmente em termos de avaliação de sistemas de RI, porque assume que a relevância refere-se, unicamente, a propriedades dos mecanismos internos do sistema.

No entanto, conforme as Figuras 5, 6 e 7 o usuário é quem decide se o objeto informacional lhe é relevante e se satisfaz sua necessidade de informação. Alguns experimentos (COOL, BELKIN e KANTOR, 1993; HOWARD, 1994; REGAZZI, 1988) apontam que os usuários utilizam outros critérios de relevância mais amplos do que os cobertos pela medida de topicalidade. Desta forma, uma visão mais extensa foi apresentada por Schamber, Eisenberg e Nilan (1990) onde a relevância é um julgamento de qualidade entre a informação recuperada e a necessidade de informação do usuário. Neste projeto será assumida esta definição e as três visões postuladas pelos mesmos autores, apresentadas no capítulo 2.

A partir de então, a relevância tornou-se um dos conceitos mais discutidos em recuperação de informação, devido à sua extrema importância para a avaliação,

^{IV} CUADRA, C.A. ; KATTER, R.V. Experimental Studies of Relevance Judgments. TM-3520/001, 002, 003/00. 3 vols., **System Development Corporation**, Santa Monica, California (1967).

desenvolvimento e projeto de sistemas de recuperação de informação (FROEHLICH, 1994). Em conseqüência, 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) e sua natureza, critérios, regiões e medidas de relevância, como identificado no capítulo 2.

Para um melhor entendimento destes aspectos do conceito de relevância, as subseções seguintes procuram esclarecê-los e discutí-los mais detalhadamente.

4.1.1 Tipos de Relevância

Na literatura, muitos pesquisadores têm adotado a tática de definir relevância usando diferentes termos que descrevem tipos distintos para esta medida. Neste projeto, será adotada a nomenclatura sugerida por Schamber (1994), em que o conceito de relevância pode ser categorizado como *orientado ao usuário* ou *orientado ao sistema*. Este autor analisou as diferentes visões de relevância que estavam sendo utilizadas na literatura e agrupou-as conforme o tipo de orientação. No Quadro a seguir, é mostrado esquematicamente os tipos de relevância.

Quadro 2 - Exemplos de tipos de relevância Fonte: MAGLAUGHLIN e SONNENWALD, 2002.

Orientada ao Sistema	Orientada ao Usuário
Lógica	Psicológica
Tópica	Situacional
Objetiva	Subjetiva

Quanto à visão orientada ao sistema, evidenciam-se três tipos, que são:

- **Lógica:** se um objeto informacional contém o assunto que se relaciona com a informação necessária em questão. Esta visão não leva em conta o julgamento original do usuário.
- **Tópica:** é baseada em suposições sobre o relacionamento entre um tópico de um documento e a pesquisa em si. Também ignora o contexto em que o usuário se insere.

^v CLEVERDON, C.; KEEN, E. Factors Determining the Performance of Indexing Systems. N. Cranfield, UK : **Aslib Cranfield Research Project**, 1966.

- Objetiva: relaciona-se com o lado procedimental do sistema, ou seja, como o sistema interpreta uma consulta efetuada pelo usuário.

Analisando estes três tipos, pode-se observar que a relevância *orientada ao sistema* preocupa-se com a correspondência ou associação entre tópico ou assunto de uma consulta e o tópico/ tema de um objeto informacional (SCHAMBER, EISENBERG e NILAN, 1990). Mais especificamente, esta orientação parte do pressuposto de que a relevância é uma propriedade do sistema e, portanto, depende de como este adquire, representa, organiza e associa as informações (SARACEVIC, 1996b).

Em suma, conforme a consideração de Park (1994), este aspecto da relevância é uma visão unidimensional dos problemas de informação dos usuários, sendo indiferente às mudanças inerentes à natureza dos problemas dos indivíduos e o subsequente impacto no processo de busca, falhando no foco do contexto complexo do indivíduo e sua tarefa. O mais agravante é que esta visão ainda é amplamente utilizada como critério de avaliação de sistemas de recuperação de informação e como um dos parâmetros para o projeto destes sistemas. Isto deve-se muito à quantidade reduzida de estudos sobre os usuários, se comparado com os estudos sobre os aspectos computacionais dos SRIs.

A partir desta problemática, atualmente a noção de relevância está se voltando em direção à natureza dinâmica que envolve a percepção do usuário, seu conhecimento e sua tarefa a resolver. Assim, surgiu a visão orientada ao usuário que, conforme o Quadro 1, é composta das seguintes dimensões:

- Psicológica: analisa os efeitos que um item pode produzir no comportamento do usuário após o julgamento feito.
- Situacional: abrange as circunstâncias que cercam as percepções dos usuários referentes à sua necessidade de informação, incluindo a situação, o contexto social, a multidimensionalidade e a dependência temporal em que os usuários estão inseridos.
- Subjetiva: o elaborador da consulta é quem efetua o julgamento referente aos itens retornados.

Como entendimento geral, a visão *orientada ao usuário* é aquela que se preocupa com o tópico do assunto e com os julgamentos dos usuários sobre os aspectos da informação. Na

verdade, envolve contextos relacionados à necessidade de informação, o propósito da mesma, como o usuário pretende encontrar a informação, sua experiência prévia no assunto de pesquisa, o estágio em que a busca pela informação se encontra, e assim por diante (PARK, 1994).

A partir das duas categorias de relevância, Saracevic (1996b) propôs um modelo denominado ‘manifestações de relevância’. Neste, o autor identifica os tipos de relevância (orientada ao sistema e ao usuário) e cita itens que servem de medidas para estes tipos de relevância. O autor também divide o processo interativo da recuperação de informação como uma ação recíproca entre os níveis do usuário (que ele denomina de cognitivo, afetivo e situacional) e o nível computacional, relacionado à engenharia, ao processamento e ao conteúdo do sistema. O Quadro 3, apresenta uma ilustração destas manifestações de relevância.

Quadro 3. Manifestações de relevância Fonte: SARACEVIC, 1996b.

Manifestações de relevância.			
Nível	Relevância	Descrição	Medidas de “sucesso”
Computacional	Sistêmica ou algorítmica	Relação entre uma consulta e objetos informacionais recuperados pelo sistema	Efetividade comparativa ou similaridades estatísticas
Computacional	Tópica	Relação entre o assunto ou tópico expressado em uma consulta e o tópico ou assunto coberto pelos objetos informacionais	Proximidade comparativa
Do usuário	Cognitiva/ pertinência	Relação entre o estado de conhecimento do usuário, sua necessidade de informação cognitiva e os objetos informacionais recuperados	Correspondência cognitiva, informatividade, novidade, qualidade da informação
Do usuário	Utilidade/ aspectos situacionais	Relação entre a situação percebida, tarefa ou problema em mãos e a utilidade do objeto de informação percebida pelo usuário	Utilidade na tomada de decisão, informação apropriada na resolução do problema, modelagem da percepção da incerteza
Do usuário	Motivacional/ afetiva	Relação entre as intenções, metas, objetivos e motivações de um usuário e os objetos informacionais recuperados pelo sistema	Sucesso, satisfação, realização

Estas manifestações de relevância são provenientes dos modelos de RI apresentados no capítulo anterior. As duas primeiras manifestações são derivadas do modelo tradicional e as restantes são derivadas conjuntamente do modelo cognitivo, do modelo baseado em eventos e do modelo baseado em camadas.

As abordagens do conceito de relevância (orientada ao sistema x orientada ao usuário) e suas manifestações de relevância tiveram como consequência direta a necessidade de rastreamento e análise dos critérios que os usuários utilizam durante o processo de RI para avaliar os objetos informacionais retornados pelo sistema. Assim, a subseção seguinte visa demonstrar e esclarecer os critérios que serão utilizados no instrumento de pesquisa deste trabalho.

4.1.2 Critérios de Relevância

Como citado anteriormente, a relevância orientada ao sistema refere-se à relação entre um objeto informacional e a consulta efetuada pelo usuário (JANES, 1994). Esta característica do processo de RI subestima a capacidade de julgamento dos usuários em relação ao sistema e aos objetos informacionais retornados, pois como visto na discussão da Figura 6, os usuários determinam a relevância dos objetos informacionais recuperados por um sistema de RI em função de uma combinação de estímulos. Estes estímulos são operacionalizados pelos usuários conforme as características dos objetos informacionais e os atributos da informação. De acordo com Bruce (1994), as características dos objetos informacionais são:

- 1 Autor (es)
- 2 Título
- 3 Descrição do assunto
- 4 Fonte de publicação
- 5 Data da publicação

O mesmo autor define os atributos da informação e alguns são citados abaixo:

- Acurácia ou exatidão: se o objeto informacional é consistente com o escopo da pesquisa efetuada.
- Conteúdo: se o conteúdo do objeto está de acordo com o assunto pesquisado.
- Sugestividade: se o objeto informacional recuperado gera novas idéias ou *insights* para o usuário.
- Completude: se o conteúdo do objeto informacional possui um boa estrutura lógica e parece dirigir-se aos aspectos críticos do problema do usuário.

A definição deste 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. Visando o atingimento de um dos objetivos desta pesquisa, definiu-se um conjunto de critérios de relevância a partir de extensa revisão bibliográfica. Estes critérios são apresentados no Quadro 4.

Quadro 4. Síntese dos critérios de relevância

Manifestação de relevância	Categorias	Crítérios	Autores
Topicalidade intelectual/ assunto	Conteúdo de informações dos OIs Intenções e metas dos usuários	Concordância tópica Informatividade do <i>abstract</i> Concordância com NI Profundidade/ escopo Percepção temporal/ atualidade	Saracevic Maglaughlin Saracevic Barry/Cool/Maglaughlin Barry/Schamber
Estrutura/ formato/	Conteúdo de informações dos OIs OIs como entidade física Crenças e preferências dos usuários	Clareza Credibilidade/validade/ qualidade/ Efetividade Facilidade de obtenção	Barry/Schamber Barry/Schamber/ Maglaughlin Barry/Schamber Barry/Schamber
Cognitiva/ pertinência	Conteúdo de informações dos OIs Experiência prévia e contexto do usuário	Aumento/ aquisição de conhecimento Novidade Novas perspectivas/ conexões Ponto de vista	Saracevic Maglaughlin Spink et al/ Cool Cool
Utilidade	Conteúdo de informações dos OIs Experiência prévia e contexto do usuário	Utilidade para NI Novidade Resolução de problema Tomada de decisão Redução de incerteza	Saracevic/Cool Saracevic/Barry/ Maglaughlin Saracevic Saracevic Saracevic
Motivacional/ afetiva	Conteúdo de informações dos OIs Experiência prévia e contexto do usuário Situação do usuário Crenças e preferências	Estímulo Novos objetivos Afeição	Bateman/Maglaughlin Spink et al/Cool/Barry Barry/Schamber/Spink et al

No Quadro anterior, os critérios estão agrupados essencialmente de acordo com as manifestações de relevância identificadas por Saracevic (1996b) (ver Quadro 3). No entanto, duas modificações foram feitas tendo em conta os objetivos. A primeira delas diz respeito à expansão do conceito de relevância tópica, de maneira a poder-se medir a percepção do usuário. Borlund (2003) denomina este novo enfoque como *topicalidade intelectual*, e explica que esta manifestação corresponde a uma avaliação intelectual do usuário sobre como um OI corresponde ao tópico/ assunto do domínio do conhecimento requerido e o descrito na busca de informação. Esta visão abranda um erro histórico de conceitualização, como discutido nas seções anteriores. Além do mais, como o foco desta pesquisa está nos aspectos adjacentes ao usuário, este novo conceito adequa-se perfeitamente aos propósitos deste projeto.

A segunda modificação refere-se à inclusão do tipo de manifestação de relevância denominada ‘estrutura ou formato’ definida por Greisdorf (2002) com a exclusão do tipo de manifestação de relevância denominada por Saracevic (1996b) de ‘algorítmica ou sistêmica’ por não ser o foco deste trabalho e por já ter sido amplamente estudada. Também no Quadro 5, pode-se perceber que alguns critérios foram identificados por mais de um autor, sendo eles Cool et al (1993), Barry (1994), Schamber (1994), Saracevic (1996b, 1997), Spink et al (1998) e Maglaughlin e Sonnenwald (2002); muitas vezes com nomenclatura diferente, porém com a mesma semântica. Uma explicação sucinta de cada critério é apresentada a seguir.

O primeiro bloco de critérios refere-se à manifestação sobre a topicalidade intelectual, sendo que os critérios possuem identificação em maior ou menor grau com o conteúdo dos objetos informacionais e as intenções e metas dos usuários. Os critérios são os seguintes:

- **Concordância tópica:** medida que refere-se a quão correta foi a interpretação dos termos da consulta efetuada pelo usuário em relação ao assunto apresentado pelo objeto informacional. Exemplos: “Este artigo não me interessa. Pesquisei sobre redes neurais e o sistema me trouxe redes de cooperação”; “Ótimo! O assunto do documento é o mesmo da minha pesquisa”.
- **Informatividade do *abstract*:** avalia o quanto a sinopse do artigo foi esclarecedora sobre o conteúdo do OI. Exemplo: “O *abstract* relata algumas coisas que não tenho muito conhecimento. Acho que vale a pena ler melhor este artigo”.
- **Concordância com NI:** mede o quanto as informações prestadas pelo OI satisfazem/ concordam com a necessidade de informação do usuário. Exemplo:

“Até que o tema do artigo corresponde as palavras que utilizei na minha busca, porém não é bem isso que estou procurando”; “Maravilha! É exatamente isto que estou procurando para meu projeto”.

- **Profundidade/ escopo:** medida que verifica o quanto o conteúdo do OI é focado no assunto de interesse do usuário e quanto o conteúdo se aprofunda no domínio do assunto. Exemplos: “Apesar de corresponder ao assunto que busquei, o artigo deixa a desejar porque a análise é muito superficial”; “Gostei do documento principalmente porque o autor faz uma descrição detalhada do problema”.
- **Percepção temporal/ atualidade:** medida que verifica se a informação provida pelo objeto é recente e atual. Exemplos: “Isto é de 2001, muito recente”; “Isto é muito antigo e já está ultrapassado”.

O segundo bloco de critérios diz respeito à manifestação de relevância sobre a estrutura/ formato. Neste bloco, permanecem alguns aspectos referentes ao conteúdo dos OIs, porém acrescidos das crenças e preferências dos usuários e, também, do artigo ou documento como uma entidade física (objeto a ser obtido) (BARRY, 1994). Os critérios são:

- **Clareza:** avalia a clareza da apresentação de um objeto informacional e a qualidade da sua redação. Exemplos: “Muitos artigos deste autor são bem escritos e editados”; “A estrutura lógica do artigo é confusa”.
- **Credibilidade/ Validade:** medida que refere-se a quão correta ou válida é a informação provida pelo objeto informacional. Exemplos: “O que o autor está dizendo é correto”; “Estes dados me parecem que são imprecisos”.
- **Efetividade:** medida para avaliação do quanto uma técnica ou procedimento apresentado no objeto informacional é efetiva ou atinge o seu objetivo com sucesso. Exemplos: “Observando os resultados, esta técnica de fato funciona”. “Devido o tamanho da amostra, a função utilizada é válida”. “Estou cético quanto a este procedimento utilizado”.
- **Facilidade de obtenção:** mede o quanto é possível ou fácil se obter uma cópia do objeto informacional. Exemplo: “Isto é uma tese de doutoramento e não foi disponibilizada para o público”.

O terceiro bloco, manifestação cognitiva/ pertinência informacional, diz respeito aos estados de conhecimento do usuário e à noção de pertinência que os objetos informacionais oferecem para a necessidade de informação dos usuários. Esta parte é acrescida da experiência, contexto e situação em que o usuário está inserido (BARRY, 1994). Os critérios identificados foram:

- **Aumento de conhecimento:** mede quanto o conteúdo provido pelo OI foi pertinente para o acréscimo de conhecimento sobre a necessidade de informação do usuário. Exemplos: “Eu sei bastante sobre a definição de risco, mas este artigo me trouxe uma nova abordagem que poderei utilizar no meu trabalho”; “Este trabalho alterou minha percepção sobre o que necessito de fato”.
- **Novidade:** mede o quanto a informação provida pelo objeto informacional é nova para o usuário. Exemplo: “Esta abordagem sobre aprendizagem organizacional eu desconhecia”; “Eu já vi esta definição muitas vezes”.
- **Novas perspectivas/ conexões:** refere-se à abertura de novas perspectivas sobre informações que possam aumentar o conhecimento do assunto pesquisado e sobre a necessidade de informação do usuário. Exemplos: “Isto me parece um bom lugar para obter novos dados e informações sobre o assunto”; “No momento, isto não é tão importante para minha dissertação mas será útil para uma próxima etapa”.
- **Ponto de vista:** medida que refere-se ao quanto as informações contidas no objeto informacional reforçam ou dão suporte ao usuário em relação ao assunto em pesquisa e sua respectiva NI. Exemplos: “Esta pesquisa fala diretamente sobre o que entendo e procuro em comunicação interpessoal”; “Este autor reforça minha visão e entendimento sobre desenvolvimento de produtos”; “Este gráfico está mostrando um efeito que eu discordo”.
- **Aquisição de conhecimento:** refere-se ao conhecimento adquirido sobre o assunto pesquisado, indo além da necessidade de informação do usuário. Exemplo: “Este artigo tirou diversas dúvidas que tinha sobre *e-learning*”; “O documento é de extrema importância pois mudou minha percepção em diversos pontos sobre flexibilidade de contratos”.

O quarto bloco de critério faz referência à manifestação sobre a utilidade do OI e os aspectos situacionais em que os usuários se encontram. Aqui, o conteúdo dos OIs e a experiência do usuário, o contexto social e suas intenções, são aspectos preponderantes durante o julgamento de relevância. Os critérios são os seguintes:

- **Utilidade para NI:** mede o quanto o objeto informacional foi útil para a necessidade de informação que o usuário possui. Exemplos: “Este artigo é de extrema importância para meu projeto atual”; “Este documento é completamente inútil para o que quero sobre cadeias de suprimento”.
- **Novidade:** mede o quanto o artigo possui novidades úteis para a necessidade de informação do usuário. Exemplo: “Muito bom este artigo! Me traz diversas novidades que poderei aplicar na previsão de demandas em serviços”.
- **Resolução de problema:** mede o quanto o OI foi útil na resolução de algum problema que o usuário possui. Exemplo: “Isso aqui vai me ajudar bastante para resolver as aplicações de passagem de parâmetros nos algoritmos”.
- **Tomada de decisão:** refere-se à utilidade do objeto informacional para ajudar o usuário a decidir sobre sua necessidade de informação. Exemplos: “Este artigo é uma boa revisão da literatura e tenho muitos artigos que foram citados aqui. Acho que já tenho o necessário para meu projeto”; “Foi bom ter lido isto pois me serviu pra mudar um pouco o foco do meu trabalho”.
- **Redução de incerteza:** mede o quanto o OI foi útil para reduzir alguma incerteza sobre o tema pesquisado e a necessidade do usuário. Exemplos: “Este artigo trouxe uma abordagem sobre modelagem de aplicações distribuídas que tirou minhas dúvidas sobre como projetar meu sistema”; “O autor descreve tão bem o comportamento do mercado financeiro que acabou com muitos questionamentos que tinha a respeito do assunto”.

O quinto e último bloco de critérios diz respeito à manifestação motivacional / afetiva. Nesta, a conjunção entre as crenças, preferências, experiências e situação do usuário são levadas em consideração durante o julgamento de relevância. Os critérios são os seguintes:

- **Estímulo:** refere-se ao estímulo que o OI ofereceu para o usuário aprofundar-se mais no assunto. Exemplo: “Isto ajuda a formular minha questão de pesquisa e procurar mais detalhes sobre o tema”.

- **Novos objetivos:** refere-se à inspiração que o conteúdo do OI fornece para que outros objetivos sejam definidos além daqueles já considerados na necessidade de informação do usuário. Exemplos: “Esta abordagem sobre marketing de relacionamento me faz pensar em ampliar meu foco de pesquisa um pouco mais do que estava pensando”.
- **Afeição:** refere-se a reações emotivas sobre alguns aspectos do objeto informacional. Nesta pesquisa este critério foi desmembrado em duas partes: 1) motivação para procurar pesquisas feitas pelo autor e 2) motivação para procurar os artigos citados pelo autor. Exemplos: “Eu adoro ler os artigos de Jarillo, eles são muito bem escritos”; “Não conheço nenhuma citação deste autor, mas talvez valha a pena pesquisar alguns”.

Acima foram mostrados os critérios a serem utilizados nesta pesquisa. Obviamente, estes não são os únicos descobertos na literatura, existindo outros, tais como influência do preço do artigo, tangibilidade, estilo de escrita, e assim por diante. No entanto, os critérios aqui identificados são suficientes para os propósitos desta pesquisa. Certamente, a depender tanto do foco da pesquisa como dos sistemas de recuperação de informação avaliados, e dos tipos de usuários pesquisados, novos e distintos critérios podem ser identificados.

Como visto até aqui, o conceito de relevância, tipos e critérios estão sendo estudados com uma maior frequência, dada a sua importância. Não obstante, faltam estudos sobre as regiões ou espaços de relevância e, em especial, sobre a correlação entre estes aspectos e os critérios (fatores de relevância). Como dito no capítulo 2, a maioria dos sistemas de RI desenvolvidos baseia-se na classificação dicotômica (relevante ou não relevante) dos objetos informacionais. Porém, alguns autores têm analisado que os usuários também julgam importantes os objetos informacionais que estejam nas regiões parciais de relevância (SPINK, GREISDORF e BATEMAN, 1998). Logo, faz-se necessário um maior esclarecimento sobre este item, que é objeto da próxima seção.

4.1.3 Regiões de Relevância

Conforme Greisdorf (2002), uma região de relevância é um domínio que serve de referência para avaliação das características dos objetos informacionais. Provavelmente, os usuários julgam estes objetos em uma região de relevância sendo, na verdade, uma classificação multidimensional, onde existem também os espaços parciais de relevância.

Pode-se imaginar, por exemplo, que um objeto informacional pode enquadrar-se no tópico do assunto requerido pelo usuário, não satisfazendo, porém, completamente sua necessidade de informação. Isto pode ser o bastante para o objeto ser julgado em algum espaço parcial de relevância. Algumas investigações feitas por Greisdorf e Spink (2000) indicam que o julgamento dos usuários, em termos destas regiões parciais, refere-se tanto ao que pode ter sido perdido no objeto informacional, como ao que tenha sido encontrado.

No entanto, objetos informacionais parcialmente relevantes podem ser potencialmente importantes para os usuários, principalmente no processo de RI que envolve um aprendizado e modificações no estado do problema. Além do mais, os critérios de julgamento de relevância fazem parte de todas as regiões de relevância, influenciando positivamente ou negativamente cada uma delas.

Infelizmente, como dito no capítulo 2, os projetos de SRI levam em consideração apenas a medida de relevância de forma dicotômica, ou seja, para um sistema de recuperação de informação o objeto informacional ou é relevante ou não. Além do mais, no momento da avaliação destes sistemas, a única dimensão que é levada em consideração é a relevância sistêmica (BORLUND, 2003). Apesar dos estudos referentes aos usuários não ignorarem a existência dos espaços parciais de relevância, eles tem tido seus dados distorcidos quando efetuada a medida de eficiência da recuperação, considerando regiões parcialmente relevantes como se fossem regiões relevantes (SARACEVIC, 1988; SU, 1994; GLUCK, 1996). Tal fato confunde os projetistas de SRI durante as etapas de confecção dos algoritmos de busca, desenvolvimento da interface e avaliação do sistema. Isto se deve principalmente ao entendimento da utilização das medidas de relevância, assunto da próxima seção.

4.1.4 Medidas de Relevância

Atualmente, não há um consenso sobre o melhor sistema para a mensuração de julgamentos de relevância e para o mapeamento de suas regiões (SPINK e GREISDORF, 2001). Isto porque a natureza dos estudos é variada. Com relação ao número de categorias, por exemplo, escalas com duas categorias de relevância (1-não relevante e 2-relevante) foram e são amplamente utilizadas. Por outro lado, Janes (1991), Pao (1993) e Borlund e Ingwersen (1997) utilizaram três categorias (1-não relevante, 2-relevância moderada e 3-relevante), Gluck (1996) utilizou duas e cinco categorias em momentos distintos, Smithson (1994) usou seis e Rees e Schultz^{VI} (*apud* Magaughlin e Sonnenwald, 2002) usaram 11.

Conforme Jacoby e Matell^{VII} (*apud* Borlund, 2003), poucas categorias de respostas resultam em uma escala muito bruta e acarreta em diminuição do poder de discriminação da taxa de resposta. Já uma escala demasiado refinada (muitas categorias) pode causar uma extrapolação na taxa de resposta.

Devido a esta conjuntura, Tang, Vevea e Shaw (1999) estudaram uma variedade de escalas de relevância, com amplitude de 2 até 11 categorias. Os resultados encontrados apontam que os níveis com pior confiabilidade estão nas escalas que possuem 2, 3 ou 11 categorias e as escalas com 5, 6 ou 7 apresentam os melhores índices de confiabilidade. Para esta pesquisa, a escala com cinco categorias (1-baixa relevância até 5-alta relevância) se mostrou a mais adequada para se atingir os objetivos propostos e isto pode ser conferido nos testes de confiabilidade, expostos no capítulo 6.

Neste capítulo, a evolução do conceito de relevância foi abordada desde os critérios focados no sistema até os direcionados para o usuário. Foram também identificados os tipos de relevância (orientada ao sistema x orientada ao usuário) e suas características. Foram indicados e exemplificados os critérios utilizados neste projeto, apresentado o que são espaços ou regiões de relevância e, por fim, analisadas as diversas medidas para o julgamento de relevância pelo usuário encontradas na literatura e definido o conjunto de medidas mais

^{VI} REES, A.M; SCHULTZ, D.G. A field experiment approach to the study of relevance assessments in relation to document searching, 2 vols. Cleveland, OH: center for Documentation and Communication Research, School of Library Science, Case Western Reserve University, 1967.

^{VII} JACOBY, J.W; MATELL, M.S. Three-point Likert scales are good enough. *Journal of Marketing Research*, 8, 495-500, 1971.

adequado para a realização desta pesquisa. De posse disto, o próximo capítulo mostra o método de pesquisa utilizado para se atingir os propósitos desta pesquisa.

5 MÉTODO DE PESQUISA

Este capítulo descreve as etapas que foram utilizadas para dar suporte à realização deste trabalho. Aqui serão apresentadas a concepção da pesquisa e as fases que compuseram os procedimentos necessários para se atingir os objetivos propostos. A seguir, serão detalhados os aspectos pertencentes ao método utilizado.

5.1 CARACTERIZAÇÃO DA PESQUISA

Dentre os diversos tipos de pesquisa, este estudo pode ser caracterizado como sendo do tipo exploratório. Conforme Mattar (1996), a pesquisa exploratória visa prover o pesquisador de maior conhecimento sobre o tema em questão, sendo apropriada para se explorar um problema ainda pouco analisado. É um tipo de pesquisa caracterizado pela flexibilidade e versatilidade quanto aos métodos empregados (MALHOTRA, 2001).

De acordo com estas definições, foram utilizadas entrevistas não estruturadas para se obter as definições das estratégias de busca dos usuários, observações foram feitas para se analisar as formulações das consultas durante o processo de RI e, por fim, foi aplicado um instrumento de pesquisa para se poder analisar os fatores de relevância empregados pelos usuários. Cada uma destas etapas será explicada nas seções subseqüentes.

Em resumo, as fases que compuseram o método da pesquisa podem ser visualizadas na Figura a seguir.

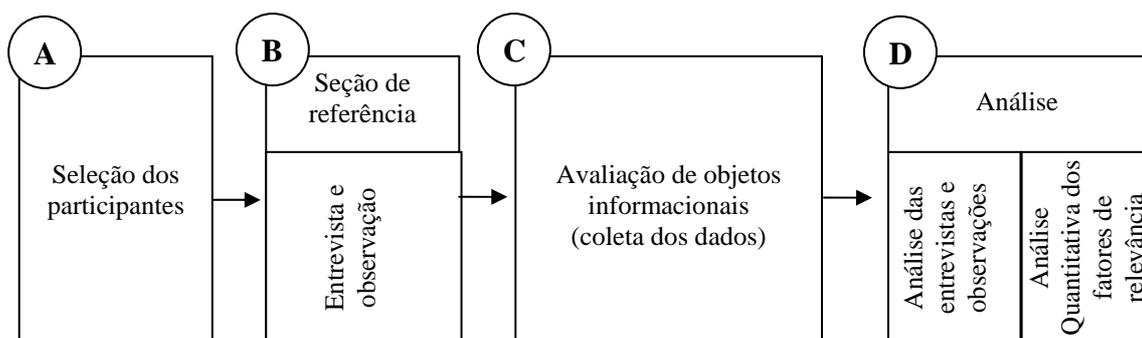


Figura 9 - Fases do método de pesquisa

5.2 SELEÇÃO DOS PARTICPANTES DA PESQUISA – FASE A

Como visto na Figura 10, a primeira etapa do desenvolvimento do método de pesquisa foi a seleção dos respondentes. Esta fase é composta por duas etapas: a população de pesquisa e a amostra. A população de pesquisa é o agregado de todos os casos que se enquadram num conjunto de especificações previamente estabelecidas (KINNEAR e TAYLOR^{VIII} *apud* MATTAR, 1996). Seguindo esta definição, a população desta pesquisa abrangeu os estudantes da UFRGS que utilizam sistemas de recuperação de informação para fornecer suporte às suas necessidades de informação.

Em seguida, foi pensado o plano amostral deste trabalho. Conforme Malhotra (2001), a amostra é um subgrupo de uma população, selecionado para participação no estudo. Nesta fase, foram considerados como elementos respondentes da pesquisa, os estudantes da UFRGS de mestrado e doutorado. A maioria destes estudantes é do PPGA (Programa de Pós-graduação em Administração) sendo os demais, participantes integrantes do PPGC (Programa de Pós-graduação em Computação) ou PPGE (Programa de Pós-graduação em Engenharia Civil), formando um total de 49 respondentes.

A decisão de se restringir a amostra à mestrandos e doutorandos deve-se, essencialmente, à característica destes usuários possuírem necessidades de informação iminentes para seus projetos, realização de artigos e trabalhos de disciplinas. Além do mais, como o objetivo primordial deste trabalho é caracterizar os fatores de relevância com um grau de riqueza acentuado, foram selecionados usuários que possuíam experiência considerável em sistemas *on-line* de RI, formulação de consultas e julgamento de objetos informacionais. Estes usuários foram contatados pessoalmente pelo pesquisador, sendo explicados a eles em que consistia o experimento e como seria realizado. Caso o candidato aceitasse, era combinado o dia, horário e local da realização do experimento. Posteriormente, o pesquisador enviava uma mensagem eletrônica para lembrar e confirmar as datas combinadas. Caso algum candidato a respondente da pesquisa ficasse impossibilitado de comparecer na data prevista ou não comparecesse sem prévio aviso, uma nova mensagem era enviada para sugerir novas opções de datas e horários.

Então, como pode ser visto, este tipo de amostra é considerada não probabilística por não ter sido utilizado nenhum conceito estatístico para sua seleção, ou seja, foram levados em

^{VIII} KINNEAR, T.C.; TAYLOR, J.R. **Marketing Research: an applied approach**. Tóquio: McGraw-Hill, 1979.

consideração os critérios subjetivos do pesquisador e a concordância com os objetivos do estudo (SAMARA e BARROS, 1997). Pode ser considerada também uma amostra por conveniência, pois sua configuração advém da facilidade de se atingir os respondentes, considerando as questões de custo, tempo e dificuldade de se atingir toda a população.

5.3 SEÇÃO DE REFERÊNCIA – FASE B

Esta fase corresponde à etapa qualitativa da pesquisa. Conforme Malhotra (2001), a fase qualitativa tem como aspecto principal o fornecimento de critérios sobre a situação-problema enfrentada pelo pesquisador e também a compreensão desta situação. Neste projeto, esta fase incluiu duas etapas: 1) revisão bibliográfica com o intuito de se elaborar de forma consistente o instrumento de coleta de dados e, 2) observações aliadas às entrevistas não-estruturadas para se capturar as definições de estratégias de busca e as formulações de consultas.

Antes dos usuários iniciarem suas buscas, era realizada uma breve entrevista para abordar quais aspectos e características os usuários utilizavam para definir suas estratégias de busca. Esta entrevista tinha um caráter informal, sendo perguntado apenas como os respondentes definiam sua busca e se possuíam algum tipo de restrição sobre os resultados. Os pontos principais das respostas eram transcritos para uma folha reservada para cada respondente. Após esta entrevista não estruturada, o respondente era convidado a realizar a busca sobre o assunto do seu interesse. O usuário tinha a liberdade de fazer quantas consultas quisesse, acessar quaisquer sistemas de recuperação de informação e poder fazer as buscas simultaneamente em vários sistemas, se assim desejasse. A única restrição era que as buscas deveriam ser feitas sobre o mesmo assunto.

Passadas estas informações, era observado como os respondentes formulavam e reformulavam suas consultas, o que os fazia modificar uma consulta e quais critérios eles utilizavam para percorrer a lista de OIs. A condução das consultas era única e exclusiva do usuário, ou seja, em nenhum momento foi feito algum tipo de intervenção ou aconselhamento por parte do pesquisador. As observações e os relatos de cada usuário também eram anotados na mesma folha citada acima.

Depois da seção de referência, era dado início ao processo de avaliação dos objetos informacionais (fase c), assunto descrito a seguir.

5.4 AVALIAÇÃO DE OBJETOS INFORMACIONAIS – FASE C

A avaliação dos objetos informacionais foi baseada na aplicação de um instrumento de pesquisa para dar suporte à coleta de dados. Então, antes de se esclarecer como se procedeu a coleta dos dados durante a seção de avaliação dos OIs, será mostrado o procedimento adaptado para se confeccionar o instrumento de pesquisa.

5.4.1 Instrumento de Pesquisa

Como dito anteriormente, o primeiro passo para se confeccionar o instrumento de pesquisa foi a revisão da literatura. Esta atividade teve como princípio o levantamento de informações, análise crítica de artigos e publicações científicas com o intuito de verificar a pertinência do projeto e, posteriormente, obter os dados necessários para a confecção do instrumento de pesquisa.

Este instrumento foi baseado em dados secundários. Para Mattar (1996), estes dados são aqueles que já foram coletados e tabulados e, tem como vantagem sua disponibilidade e seu baixo custo. Normalmente, as fontes básicas são publicações, governos, instituições, entre outros. No caso deste projeto, os resultados desta etapa qualitativa vieram preponderantemente dos artigos confeccionados pelas autoridades que estudam o conceito de relevância. Estes dados referem-se aos critérios de relevância que foram discutidos no capítulo 4.

De posse dos dados, deu-se início à confecção e estruturação do questionário. Segundo Malhotra (2001), o questionário é uma técnica estruturada para coleta de dados, que consiste de uma série de perguntas - escritas ou verbais - que um entrevistado deve responder. A estrutura do questionário utilizado nesta pesquisa é composta de três blocos (ver Anexo A). O primeiro deles visa identificar o perfil do respondente e o último bloco coleta algumas informações pessoais do usuário para fins de controle.

O segundo e mais importante bloco, refere-se à classificação de relevância que cada critério pode obter. O questionário foi composto por 5 dimensões que perfizeram 23 critérios de relevância a serem mensurados (Quadro 4, capítulo 4). Foi utilizada a escala *Likert* de cinco pontos para avaliação dos critérios e também para julgamento da manifestação de relevância como um todo. Como dito na seção referente à medidas de relevância, a escala de

cinco pontos adequa-se a este projeto por demonstrar resultados empíricos satisfatórios e por esta quantidade de categorias ser condizente com os propósitos deste trabalho.

Vale a pena ressaltar que antes de se obter a versão final do instrumento, dois pré-testes foram conduzidos com os seguintes objetivos:

- avaliar a compreensão das questões por parte dos respondentes;
- verificar a adequação da escala;
- avaliar a interface gráfica;
- medir o tempo médio que durava o experimento por completo.

O primeiro pré-teste foi realizado com seis pessoas. Segundo Malhotra (2001), a melhor forma de se efetuar o pré-teste é com entrevistas pessoais porque os entrevistadores podem observar as reações e atitudes dos entrevistados. Do resultado do primeiro pré-teste, houve uma necessidade de se ajustar alguns termos dos questionamentos dos critérios e modificar o *layout* do questionário. Assim, palavras foram modificadas ou substituídas por sinônimos e os critérios foram agrupados em blocos.

O segundo pré-teste resultou apenas na alteração da posição de alguns dos critérios. Neste teste, os respondentes não tiveram dificuldade em assimilar os questionamentos dos critérios. O tempo do experimento variava de acordo com a quantidade de consultas realizadas e de objetos informacionais julgados e, em média, o tempo gasto foi de uma hora e 18 minutos. Por fim, vale ressaltar que cada experimento foi realizado utilizando um respondente por vez.

A partir disto, deu-se a fase quantitativa da pesquisa, envolvendo os procedimentos de coleta e processamento dos dados, assuntos abordados a seguir.

5.4.2 Procedimento de Coleta dos Dados

Conforme Dillon, Madden e Firtle (1987), a fase quantitativa da coleta dos dados é a mais importante etapa isolada da pesquisa, pois ainda que a definição e estruturação do estudo sejam importantes para a solução do problema, qualquer aspecto deficiente na obtenção quantitativa dos dados pode invalidar as informações e distorcer as futuras conclusões.

Levando em conta estas considerações, algumas decisões foram tomadas para se otimizar o processo de coleta dos dados. Inicialmente, a escolha dos OIs era uma decisão

unicamente do usuário, de forma que o pesquisador não exercia nenhuma interferência nesta decisão.

Não era imposta uma quantidade mínima ou máxima de OIs a serem recuperados e julgados pelos respondentes (a média dos OIs julgados foi de 4,85). Após a consulta efetuada, os respondentes tinham total liberdade para selecionar os documentos que seriam por eles analisados e poderiam reescrever, expandir ou limitar suas consultas e, como dito anteriormente, a única restrição era que estas novas consultas abordassem o mesmo assunto descrito pelos participantes da pesquisa. Ou seja, todas as buscas deveriam obedecer ao mesmo assunto declarado pelo respondente (questão 1, anexo A)

Para cada OI selecionado havia um questionário a ser respondido, sendo que cada um destes questionários preenchidos já permanecia nas mãos do pesquisador (resposta dos objetos informacionais *ex ante*). Cada objeto analisado era impresso e a ele, o mesmo tipo de questionário era anexado. Também nesta etapa eram anotadas algumas opiniões que os usuários atribuíam aos objetos informacionais julgados, sendo que muitas vezes estas opiniões representavam novos critérios de relevância.

Finalizada a etapa de julgamento dos objetos informacionais *ex ante*, um prazo de devolução dos questionários respondidos (avaliação dos OIs *ex post*) era combinado com o respondente. Para isto, era levado em conta a quantidade de OIs analisados^{IX}. Feito isto, era enviado um *e-mail* inicial para servir como lembrete do prazo estabelecido. Caso o respondente não cumprisse o prazo por algum motivo, uma nova data era estabelecida.

Após esta importante etapa, deu-se início à etapa de processamento dos dados e assim, obtenção dos fatores de relevância.

^{IX} A média de devolução foi de 23 dias.

5.5 ANÁLISE – FASE D

Como visto na Figura 10, esta fase é composta por duas partes: a análise das entrevistas e observações e a análise quantitativa dos fatores de relevância. A primeira delas baseou-se nas anotações feitas durante o processo de recuperação de informação. Estas anotações eram avaliadas para se extrair e agrupar as estratégias de busca, táticas de consulta e novos critérios que eram citados de formas distintas, mas possuíam a mesma semântica. A segunda fase teve um aspecto mais quantitativo pela necessidade de se processar os dados do instrumento de pesquisa e pela utilização de técnicas estatísticas. De acordo com Mattar (1996), o processamento dos dados compreende os passos necessários para transformar os dados brutos coletados em dados trabalhados que permitirão a realização das análises e interpretações.

A primeira tarefa é a verificação dos dados no intuito de impor um padrão mínimo de qualidade aos dados brutos, otimizando sua precisão e minimizando as ambigüidades. (MATTAR, 1996). Nesta pesquisa, esta fase envolveu a inspeção dos dados de cada instrumento de pesquisa coletado, sendo verificado se ele estava completo, consistente e preciso. A segunda etapa do processamento diz respeito à codificação dos dados (procedimento técnico em que os dados são categorizados). Ou seja, transformação destes dados em símbolos numéricos para que pudessem ser tabulados e contados através do processamento. Neste caso, apenas as respostas referente às bases de dados foram codificadas, já que os demais estavam corretos para fins de análise estatística.

Posteriormente, os dados dos questionários foram digitados na planilha eletrônica Excel[®] e depois importados para o *software* estatístico *Statistical Package for the Social Science* - SPSS[®] versão 11.0. Como houve esta transferência de dados entre programas de computadores, uma nova depuração dos dados foi realizada com o objetivo de verificar se a consistência dos dados foi mantida, sendo exercida uma comparação entre as bases de dados. Finalizada esta fase, deu-se início ao tratamento estatístico dos dados para se gerar os resultados.

Inicialmente, foi testada a confiabilidade interna do instrumento de coleta dos dados tendo como parâmetro o *Alpha de Cronbach*. A seguir, utilizou-se a estatística descritiva para melhor visualização da composição da amostra e análise das médias de relevância. Por fim, foi efetuado o teste estatístico mais importante para se atingir o objetivo principal – a

regressão linear múltipla. O capítulo seguinte discute a obtenção destes resultados, bem como elucida as técnicas estatísticas utilizadas.

6 RESULTADOS DA PESQUISA

Neste capítulo são apresentados os resultados do presente estudo. Estes resultados estão divididos em duas etapas, uma etapa exploratória qualitativa e outra exploratória quantitativa. A primeira delas explica os resultados das fases de definição da estratégia de busca e da formulação da consulta. Na segunda etapa, inicialmente serão visualizados os testes de confiabilidade interna do instrumento de pesquisa. Posteriormente, os dados de caracterização da amostra serão descritos, sendo analisadas as variáveis referentes à finalidade da necessidade de informação, à atividade acadêmica exercida pelo respondente, nível de conhecimento em relação ao tema pesquisado, os *sites* de busca mais utilizados pelos respondentes e a distribuição das respostas dos espaços de relevância.

Em seguida, é demonstrado o resultado da análise univariada, expondo o comportamento das respostas de cada variável nos dois momentos de julgamento de relevância (*ex ante* e *ex post*). Na verdade, esta etapa mostra uma análise preliminar dos critérios de relevância, sendo indicadas as medidas de tendência central e de dispersão.

A última parte da análise diz respeito à utilização da técnica multivariada. Mais especificamente, a técnica utilizada é a regressão linear pois, conforme Mattar (1996), a regressão refere-se à natureza da associação estatística, ou seja, à correspondência de uma variável critério em relação a uma ou mais variáveis prognóstico. Portanto, esta técnica é de extrema importância para um dos objetivos desta pesquisa, devida à possibilidade de se avaliar a influência dos critérios em cada espaço de relevância (caracterização dos fatores de relevância).

6.1 ETAPA EXPLORATÓRIA QUALITATIVA

6.1.1 Definição das Estratégias de Busca

Esta é a primeira fase durante o processo de definição dos fatores de relevância, conforme a Figura 1. Sua importância para os projetos de sistemas de recuperação de informação está no fato de entender as táticas que os usuários utilizam para representar suas necessidades de informação. Nesta pesquisa foram identificadas quatro estratégias de busca relatadas pelos usuários. A depender do participante da pesquisa, estas estratégias eram utilizadas isoladamente (enfoque apenas em uma estratégia) ou simultaneamente (utilização conjunta de mais de uma estratégia). A seguir está a identificação de cada uma delas:

1. Abrangência do tema em pesquisa: esta estratégia era utilizada por todos os participantes. Inicialmente, era relatada a necessidade de informação (questão 1 do anexo A), o problema que o usuário possuía e o conhecimento prévio sobre o tema abordado. Desta conjuntura derivavam-se as palavras-chave que, normalmente, incluíam alguns termos da NI. Exemplo citado por um dos respondentes: “Eu quero artigos que envolvam a sumarização e técnicas de inteligência artificial” Palavras-chave: *summarization, artificial, intelligence*.
2. Afetividade com autor: esta segunda estratégia era bastante utilizada pelos usuários avançados e médios, mas alguns usuários iniciantes também fizeram uso desta. Normalmente, era identificado o autor de afinidade do usuário e proveniente disto, as palavras-chaves eram compostas pelo nome do mesmo. Exemplo citado por respondente: “A maioria de minhas buscas é feita usando algum autor conhecido tipo Gulati ou Arino”. Palavra-chave: Gulati.
3. Afinidade com a fonte: estratégia praticamente utilizada pelos usuários avançados e médios. Decorrente do conhecimento prévio sobre o assunto, os usuários acessavam os periódicos validados por eles e percorriam suas edições. Aqui, a utilização de palavras-chave dependia da base de busca, já que a maioria disponibilizava um catálogo de quais periódicos eram indexados.
4. Busca por citação de autor reconhecido: esta estratégia também era muito utilizada pelos usuários médios e avançados. Neste caso, o usuário procuraria

os artigos citados em trabalhos realizados pelo autor que ele reconhece como especialista na área.

Normalmente, as mesclas das estratégias eram feitas unindo o nome do autor e o assunto de interesse do usuário. Outro aspecto importante das estratégias de busca era a utilização do ano como fator delimitador. Dos 49 respondentes, apenas dois demonstraram não se importar com a data de publicação do objeto informacional. Para os demais, o ano funcionava como um ponto de corte para a consulta.

Depois de identificadas as estratégias, foram observadas a formulação da consulta e a seleção dos OIs de interesse; ambos discutidos na próxima seção.

6.1.2 Formulação das Consultas e Seleção Inicial dos OIs

Esta etapa é proveniente das palavras-chave derivadas das estratégias de busca. Primeiramente, foi observado que na maioria das vezes, os usuários praticamente não utilizavam os recursos avançados de busca disponibilizados pelos sistemas, se resumindo apenas aos conectivos *and* e *or*.

De forma sucinta, as táticas de busca poderiam ser assim identificadas:

1. *Top-down*: aqui, os usuários utilizavam termos mais abrangentes do assunto em pesquisa. Feito isto, rastreavam a lista dos objetos informacionais recuperados pelo sistema e aprendiam novos termos ou termos adjacentes sobre o assunto e utilizavam-nos para a reformulação das consultas. Sucintamente, esta era uma tática utilizada pelos usuários iniciantes e alguns médios, envolvendo um processo de abordagem e aprendizagem sobre o assunto e incluía refinamentos sucessivos da consulta.
2. *Middle*: tática utilizada pelos respondentes considerados médios e por alguns avançados. Aqui, os usuários mesclavam termos abrangentes com alguns específicos do assunto e, conforme os resultados da consulta, eles tomavam as seguintes direções: A) se a quantidade de objetos informacionais recuperados fosse pequena, os usuários expandiam a busca retirando palavras ou substituindo-as por termos mais genéricos para capturar novos OIs; B) se a lista de objetos informacionais fosse muito grande os usuários restringiam mais ainda a busca, acrescentando novos termos.

3. *Bottom-up*: tática utilizada exclusivamente pelos usuários avançados. As consultas possuíam termos extremamente específicos da área de interesse e utilização em excesso dos conectivos *and*. Caso não retornasse nenhum objeto informacional ou a lista recuperada fosse muito pequena, alguns termos eram retirados ou trocados por sinônimos até se obter uma quantidade razoável de OIs.

Na maioria das consultas existia uma seleção de alguns OIs que aparentassem ser de interesse do usuário. Pelo que foi observado, esta seleção era feita em uma seqüência de passos bem definida. Inicialmente, os usuários se interessavam pelos títulos dos objetos informacionais que contivessem todos ou alguns termos de suas consultas. Caso isto não acontecesse ou os títulos apresentassem alguns termos que abordasse outra área de conhecimento, o OI era descartado. Senão, os respondentes liam o *abstract* para verificar se o artigo descrevia o assunto de interesse ou não. Caso representasse referência sobre o assunto abordado pelo usuário, o conteúdo do OI era analisado para se verificar a correspondência com a necessidade de informação do usuário.

Foi observado durante a análise *ex ante* e também descrito pelos usuários na análise *ex post*, que para algumas partes do objeto era dada uma ênfase maior como, por exemplo, nos resultados. Além do mais, os respondentes citaram que algumas vezes, certos critérios eram utilizados em mais de uma ocasião para se julgar o objeto informacional. Isto foi identificado nos resultados da análise multivariada, onde serão discutidas mais detalhadamente estas características.

Por fim, da definição da estratégia de busca, seleção dos OIs de interesse e das análises *ex ante* e *ex post*, os respondentes citaram alguns critérios de relevância que não estavam especificados no instrumento de pesquisa. A próxima seção mostra e explica estes critérios.

6.1.3 Critérios Indicados pelos Respondentes

Abaixo estão elencados os critérios citados pelos respondentes da pesquisa durante o processo de RI:

- Artigos de fonte de publicação reconhecida internacionalmente
- Credibilidade da Universidade onde o autor do artigo trabalha
- Ano de publicação do artigo

- Artigos de mercados emergentes
- Número de páginas do objeto informacional

O primeiro critério identificado é proveniente da estratégia de busca do usuário e foi citado por 28 respondentes dos 49 desta pesquisa. Se a fonte de publicação não for reconhecida internacionalmente quanto à sua qualidade e seriedade, os artigos contidos nela são praticamente descartados. Caso contrário, o usuário se interessa em percorrer os artigos publicados neste periódico.

O segundo critério, credibilidade da universidade onde o autor do artigo exerce suas atividades profissionais, diz respeito à validação e imagem que o usuário atribui à excelência da instituição. Este critério parece ser um complemento ou alternativa da definição da estratégia baseada na afetividade com o autor. Ou seja, se o OI parece ser interessante mas não é de um autor reconhecido na área, o próximo atributo a ser analisado é a instituição em que esse autor exerce suas atividades de pesquisa.

O terceiro critério também é decorrente da estratégia de busca. Quase todos os respondentes se importavam com a idade do objeto informacional. Mais especificamente, os artigos recuperados deveriam ter data de publicação a partir da década de 90. As únicas exceções diziam respeito a artigos considerados marcos do assunto.

O quarto critério foi pouco citado pelos respondentes, mas sua identificação pode ter implicações para futuras pesquisas. Alguns usuários declararam que artigo cujo conteúdo refere-se à aplicações e abordagens de mercados e países emergentes, trazia visões interessantes e muito diferentes do convencional, o que poderia ser útil para suas pesquisas.

O último critério, número de páginas do artigo, também teve uma boa quantidade de citação pelos respondentes. O argumento utilizado é que artigos com poucas páginas, normalmente é muito superficial. Perguntado então, qual seria o mínimo de páginas a ser considerado, as respostas variavam entre seis e nove páginas.

Em suma, a definição da estratégia, a formulação da consulta e estes critérios de relevância estão bem acoplados uns aos outros. Após estes resultados provenientes das entrevistas não estruturadas e observações do comportamento do usuário durante o processo de RI, foi analisada a etapa quantitativa desta pesquisa, assunto das próximas seções.

6.2 ETAPA EXPLORATÓRIA QUANTITATIVA

Esta etapa aborda todo processo quantitativo necessário para se atingir o objetivo desta pesquisa.

6.2.1 Teste de Consistência Interna

Com o objetivo de analisar a consistência interna de cada uma das cinco manifestações de relevância, foi efetuado o teste denominado *Alpha de Cronbach*. Conforme Dillon, Madden e Firtle (1987), este teste é a forma mais aceitável e utilizada para os testes de coerência interna. Este coeficiente varia de 0 (zero) a 1 (um), e um valor de 0,6 ou menos geralmente indica confiabilidade insatisfatória da consistência interna (MALHOTRA, 2001).

Outro aspecto importante deste teste para esta pesquisa, deve-se principalmente à montagem do instrumento de coleta dos dados. Como visto no capítulo 4, este instrumento foi criado com base na teoria de manifestações de relevância; sendo que cada critério que compunha estas diversas manifestações foi levantado em estudos de diferentes autores e em épocas distintas (ver quadro 4). Disto, advém a necessidade de se testar a confiabilidade interna de cada manifestação. A Tabela 1 mostra os dados deste teste.

Tabela 1 - Testes de confiabilidade

Manifestação	Valor do <i>Alpha</i> (<i>ex ante</i>)	Valor do <i>Alpha</i> (<i>ex post</i>)
Topicalidade Intelectual	0,7792	0,7964
Estrutura/ formato	0,7039	0,7078
Cognitiva/ pertinência	0,9068	0,9425
Utilidade	0,9280	0,9543
Motivacional	0,8110	0,8842

Como pode ser visto, todos os blocos referentes às manifestações de relevância apresentaram índices de confiabilidade acima do exigido, autenticando o instrumento para ser utilizado neste estudo.

6.2.2 Caracterização da Amostra

Como pode ser visto na Tabela abaixo, a amostra é composta por 49 respondentes que julgaram no primeiro momento 238 objetos informacionais e, desta coleta, foram retornados 224 OIs; o que caracteriza uma taxa de retorno de 94,21% dos objetos julgados.

Tabela 2 - Composição da amostra

Respondentes	Quantidade <i>ex ante</i>	Quantidade <i>ex post</i>
49	238	224
Taxa de retorno: 94,12%		

6.2.3 Atividade Acadêmica

Os resultados da Tabela 3 mostram que a amostra era composta na sua maioria por estudantes de mestrado (67,3 %). Isto deve-se, principalmente, por se tratar de uma amostra por conveniência, em que a quantidade de respondentes ao alcance do pesquisador era deste perfil.

Tabela 3 - Atividade acadêmica

Atividade Acadêmica	Frequência	Percentual
Mestrado	33	67,351
Doutorado	16	32,649
Total	49	100,0

6.2.4 Finalidade

Estes resultados mostram os objetivos distintos a que se propunha o uso dos objetos informacionais.

Tabela 4 - Finalidade da informação

Finalidade	Frequência	Percentual
Artigo	3	6,1
Dissertação	30	61,2
Tese	16	32,7
Total	49	100,0

Como visto acima, o desenvolvimento da dissertação (61,2%) é a finalidade mais utilizada nesta amostra. Em segundo vem a utilização para a tese e, por último, o uso para se escrever artigos. Estes resultados entram em concordância com os dados identificados na Tabela 3, já que a maioria é de mestrados e, portanto, em fase de confecção dos projetos de dissertação.

6.2.5 Nível de Conhecimento

No que se refere ao nível de conhecimento dos respondentes em relação ao tema por eles pesquisados, a Tabela 5 mostra os seguintes dados.

Tabela 5 - Nível de conhecimento por atividade acadêmica

Atividade Acadêmica	Nível	Frequência	Percentual
Mestrado	Iniciante	8	16,33
	Médio	12	24,49
	Avançado	13	26,53
	Total	33	67,35
Doutorado	Iniciante	4	8,17
	Médio	5	10,20
	Avançado	7	14,28
	Total	16	32,649
Total		49	100

Pode-se perceber que para ambos os perfis dos respondentes (mestrados ou doutorandos), existia uma quantidade maior de usuários avançados, seguido dos médios e, por

último, os considerandos iniciantes. No entanto, a diferença numérica não era muito acentuada de um nível para outro.

6.2.6 Bases de Busca Utilizadas

A Tabela 6 mostra os dados referentes às quantidades de OIs recuperados em cada base de busca.

Tabela 6 - Bases de busca utilizadas

Finalidade	Frequência	Percentual
Science Direct	109	45,8
Proquest	45	18,9
EBSCO	40	16,8
Google	4	1,7
MSN	6	2,5
SSRN	22	9,2
ACM	10	4,2
CFPM	2	0,8
Total	238	100,0

Como pode ser visto, *Science Direct* foi a base que teve a maior frequência de objetos recuperados (45,8%), seguida pela *Proquest* (18,9%) e pela EBSCO (16,8%). A base ACM (4,2%) foi utilizada unicamente pelos respondentes integrantes do PPGC.

Pode ser visto também que *sites* de busca não especializados em artigos científicos foram utilizados, tais como *Google* (1,7 %) e *MSN* (2,5 %).

6.2.7 Distribuição das Respostas

Tabela 7 - Distribuição das respostas de relevância

Valor de Relevância	Ex ante		Ex post	
	Frequência	Percentual	Frequência	Percentual
1	9	3,8	15	6,7
2	37	15,5	45	20,1
3	60	25,2	52	23,2
4	93	39,1	72	32,1
5	39	16,4	40	17,9
TOTAL	238	100,0	224	100,0

A Tabela acima mostra a distribuição dos julgamentos dos OIs em termos de relevância, para os dois momentos de medidas. Pode-se notar que, inicialmente, 93 OIs foram

julgados como relevantes (valor 4) e esta quantidade decresceu para 72 (*ex post*). De uma forma geral, as menores quantidades de classificação de OIs estão nos extremos da relevância (valore 1 e 5).

6.2.8 Análise Univariada

A Tabela 8 mostra o comportamento do bloco topicalidade intelectual e dos seus critérios individualmente.

Tabela 8 - Comportamento geral dos critérios de topicalidade intelectual

Topicalidade Intelectual/ Assunto	Média (<i>Ex Ante</i>)	D.P.	Média (<i>Ex Post</i>)	D.P.
1. Concordância entre os termos da consulta efetuada e o assunto apresentado pelo documento.	4,16	1,02	3,99	1,06
2. Informatividade do <i>abstract</i> para sua decisão de aprofundar-se no texto completo.	4,10	1,00	3,89	1,08
3. Concordância entre as informações apresentadas no documento e sua necessidade de informação.	3,59	1,15	3,41	1,18
4. Profundidade/ foco do conteúdo no documento.	3,82	1,05	3,68	1,11
5. Atualidade do assunto apresentado.	4,23	0,88	4,20	0,98
TOTAL	3,92	0,96	3,78	1,01

Na etapa *ex ante* dos julgamentos, esta manifestação apresentou uma média (3,92) de relevância sendo levemente superior à *ex post* (3,78). Em termos dos critérios individualmente, esta tendência de queda da média de julgamento foi mantida. Pode ser visto também que a atualidade do assunto apresentado (critério 5) foi o que apresentou maiores índices de relevância, situando-se entre os espaços considerados relevante (valor 4) e altamente relevante (valor 5). Esta característica entra em concordância com os dados da definição da estratégia de busca, já que um dos aspectos importantes para se planejar a busca era a necessidade do conteúdo e do OI em si serem recentes e atuais. Obviamente, os artigos analisados possuíam estes atributos e, portanto, tiveram altas classificações de relevância.

A concordância entre os termos da consulta efetuada e o assunto apresentado pelo objeto informacional (critério 1) também obteve bons índices de relevância. O que caracteriza que boa parte dos OIs selecionados atendia à consulta especificada pelo usuário. Já o critério 3, concordância entre as informações apresentadas no documento e a necessidade de

informação do usuário, foi a variável que apresentou as menores médias de relevância. Os critérios 2 e 4 possuíram médias que se situaram no espaço considerado relevante (valor 4).

Abaixo podem ser visualizados os dados referentes à manifestação estrutura/ formato dos objetos informacionais.

Tabela 9 - Comportamento geral dos critérios da manifestação de estrutura/ formato

Estrutura	Média (Ex Ante)	D.P.	Média (Ex Post)	D.P.
7. Clareza apresentada no documento.	3,89	0,94	3,97	0,91
8. Credibilidade (informações corretas e válidas) provida pelo documento.	3,97	0,98	3,99	0,99
9. Efetividade do método ou procedimento utilizado pelo autor.	3,66	1,01	3,69	1,11
10. Facilidade de obtenção do documento.	4,54	0,74	4,57	0,73
TOTAL	3,89	0,85	3,89	0,91

O comportamento desta manifestação apresentou médias de relevância acima do moderado (valor 3) e, casualmente, a média total deste bloco permaneceu idêntica nas duas etapas de medição. No entanto, observando par a par as medidas dos critérios, evidencia-se um pequeno aumento das médias de relevância na última tomada de medição (*ex post*).

A facilidade de obtenção do documento (critério 10) foi o fator que obteve a maior média de relevância nos dois momentos do estudo. Obviamente, isto é confirmado nas definições da estratégia de busca. Os usuários normalmente fazem busca em sistemas que oferecem acesso completo ao objeto. A dificuldade de obtenção é considerada quando existe um custo financeiro para se obter o objeto ou quando diversas consultas foram efetuadas em tempos distintos para obtê-lo. Os demais critérios mantiveram-se com médias entre as regiões moderada (valor 3) e relevante (valor 4). A credibilidade e a efetividade tiveram mudanças de valores bem pequenas entre as duas medições. Isto é proveniente do enfoque demasiado que os usuários davam à estes atributos informacionais, tanto no momento *ex ante* quanto no *ex post*.

No que se refere à clareza estrutural e de conteúdo do documento, verifica-se que as médias tiveram uma modificação mais acentuada do que os critérios anteriores. Certamente, esta mudança deve-se ao processo de seleção dos OIs. No primeiro momento, o usuário possui uma visão geral do documento e o analisa em relação à estruturação das seções do documento e quanto à facilidade de leitura, descartando os OIs que possuem linguagem e

estruturação confusas. Assim, quando a leitura completa é exercida, a tendência é de se verificar e até melhorar a percepção em relação à clareza do documento.

A Tabela 10 diz respeito ao comportamento da tendência central da manifestação de relevância denominada cognitiva ou pertinência informacional, bem como as avaliações de cada critério.

Tabela 10 - Comportamento geral dos critérios da manifestação cognitiva

Cognitiva/ Pertinência	Média (Ex Ante)	D.P.	Média (Ex Post)	D.P
12. Aumento de conhecimento acerca de sua necessidade de informação.	3,69	1,07	3,41	1,19
13. Obtenção de novas informações sobre o assunto pesquisado.	3,86	1,04	3,60	1,24
14. Ampliação de novas perspectivas de conhecimento.	3,72	1,17	3,53	1,29
15. Conhecimento adquirido sobre o assunto pesquisado.	3,47	1,07	3,35	1,20
16. Reforço ou suporte ao seu ponto de vista sobre o assunto pesquisado.	3,50	1,16	3,31	1,28
TOTAL	3,66	1,03	3,46	1,19

Pode-se perceber que este bloco praticamente manifestou uma tendência homogênea de decréscimo das médias em relação às duas medições. Inicialmente, a manifestação como um todo apresentou uma queda nos seus valores de média (de 3,66 para 3,46) e, esta tendência foi mantida internamente pelos critérios, no que diz respeito à inserção dos valores próximos do espaço moderado de relevância (entre 3 e 4).

A obtenção de novas informações sobre o assunto pesquisado (critério 13) foi o critério que apresentou os maiores índices de tendência central tanto na primeira (3,86) quanto na segunda medida (3,60). Levando-se em consideração que a quantidade de iniciantes acrescida dos considerados médios indicam quase 58% da amostra, é provável que a obtenção de novas informações seja algo mais fácil para estes perfis de usuários e que, a frustração também seja acentuada após a leitura completa do documento devido à constante mudança de foco.

O segundo maior valor de média foi assumido pelo critério referente à ampliação de novas perspectivas de conhecimento (critério 14). Vale a pena ressaltar que o conhecimento adquirido sobre o assunto (critério 15) foi o critério que apresentou a menor média (3,47) no

primeiro estágio de medida. Posteriormente, na medida *ex post*, a menor média foi assumida pelo critério 16, reforço ou suporte ao ponto de vista do usuário.

A Tabela 11 mostra os dados referentes à manifestação de utilidade do objeto informacional. Como pode ser percebido, os julgamentos feitos neste bloco tiveram uma tendência mais próxima da relevância moderada à parcialmente não relevante na medida *ex post* (valores entre 2 e 3). Os dados também mantiveram a tendência de queda dos valores em relação ao primeiro momento de medida.

Tabela 11 - Comportamento geral dos critérios de utilidade

Utilidade	Média (Ex Ante)	D.P.	Média (Ex Post)	D.P.
18. Utilidade para sua necessidade de informação.	3,57	1,21	3,31	1,34
19. Apresentação de novidades úteis para sua necessidade de informação.	3,51	1,16	3,29	1,35
20. Utilidade na resolução do problema informacional que você tem em mãos.	3,14	1,18	2,87	1,28
21. Utilidade para tomar uma decisão sobre a necessidade de informação necessária.	3,13	1,19	2,95	1,28
22. Utilidade para reduzir alguma incerteza quanto à sua necessidade de informação.	3,07	1,25	2,92	1,30
TOTAL	3,31	1,15	3,12	1,28

A utilidade do objeto informacional analisado (critério 18) apresentou as maiores médias de relevância tanto no momento *ex ante* quanto no *ex post*. Já a utilidade para redução de alguma incerteza informacional (critério 22) apresentou a menor média de julgamento no momento *ex ante*. No entanto, a menor média assumida no momento *ex post* passou a ser do critério 20.

De forma geral, dos cinco critérios que fazem parte desta manifestação, três decresceram suas médias consideravelmente, sendo que a utilidade na resolução de algum problema informacional foi o critério que apresentou queda mais drástica na média de avaliação (cerca de 0,27 pontos percentuais).

Como a medida *ex ante* baseia-se na primeira impressão de utilidade que o OI transmite, é esperado que com a leitura mais aprofundada, esta utilidade tenha um decréscimo devido à maior percepção do conteúdo do documento. Além do mais, a conjuntura entre usuários de médio conhecimento com os avançados permite maior poder de decisão sobre o que é útil ou não para suas finalidades, após uma leitura mais crítica.

Já a Tabela 12 mostra os dados referentes à manifestação motivacional. Mais uma vez, todas as médias apresentaram decréscimo de valor em relação à primeira tomada de medida.

Tabela 12 - Comportamento geral dos critérios da manifestação motivacional

Motivacional	Média (Ex Ante)	D.P.	Média (Ex Post)	D.P.
24. Estímulo que o conteúdo do documento lhe ofereceu a aprofundar-se mais no assunto.	3,51	1,21	3,26	1,25
25. Inspiração a pesquisar outros objetivos além de sua necessidade de informação.	3,09	1,26	3,01	1,27
26. Motivação a procurar pesquisas feitas pelo mesmo autor.	2,80	1,28	2,56	1,32
27. Motivação a procurar os artigos citados pelo autor.	3,23	1,32	3,04	1,41
TOTAL	3,28	1,12	3,09	1,23

Um dado interessante é que a única variável que não atingiu o espaço de relevância considerado moderado foi a motivação a procurar pesquisas feitas pelo mesmo autor (critério 26). Já o critério referente ao estímulo que o conteúdo do objeto oferece para o usuário aprofundar-se mais no assunto em pesquisa (questão 24), foi o que obteve melhor avaliação da média de relevância em ambas as medidas. Os demais critérios mantiveram-se próximos do espaço moderado de relevância.

Concluindo esta etapa de análise univariada, ela é importante para se ter uma idéia geral dos julgamentos de relevância de cada manifestação e de cada critério, podendo ser entendida como uma extensão da caracterização da amostra. Porém, não se pode fazer inferências sobre a associação entre os critérios, manifestações de relevância e o julgamento do OI como um todo. Para haver esta possibilidade, fez-se uso da regressão linear múltipla, que é um teste estatístico mais avançado e será explicado na seção seguinte.

6.2.9 Análise Multivariada

Como dito no início deste capítulo, a análise multivariada foi focada na técnica denominada, análise de regressão. Conforme Malhotra (2001), a análise de regressão é um processo poderoso e flexível para a análise de relações associativas entre uma variável métrica dependente e uma ou mais variáveis independentes. De acordo com o mesmo autor, as variáveis independentes são variáveis ou alternativas que são manipuladas pelo pesquisador e cujos efeitos são medidos e comparados. Já as variáveis dependentes medem os efeitos das variáveis independentes sobre as unidades de teste.

No caso deste trabalho, o primeiro tratamento estatístico visou avaliar a influência dos critérios de relevância para o julgamento completo do OI, com o objetivo de se explorar e entender o grau de importância que cada critério exerce nos espaços de relevância (Tabelas 13, 14, 15, 16 e 17). Assim, cada critério foi considerado uma variável independente e a avaliação de relevância do documento (ver questão 29, anexo A) era considerada a variável dependente. No segundo tratamento estatístico (Tabela 18), foi feita uma regressão linear para se identificar a influência das manifestações de relevância no julgamento completo do objeto informacional. Ou seja, nesta etapa as variáveis independentes passaram a ser as questões de fechamento de cada bloco de manifestação (questões 6, 11, 17, 23 e 28 do instrumento de pesquisa) e a variável dependente era a mesma do primeiro tratamento (questão 29).

Porém, antes de analisar os resultados obtidos, é oportuno descrever os componentes das tabelas. Nas cinco primeiras Tabelas (13, 14, 15, 16 e 17), existem as identificações dos critérios analisados, os valores dos julgamentos de relevância (1-baixa relevância até 5-alta relevância) e a ordem de influência de cada critério em cada uma das respostas de relevância. Para cada uma das medições, existem quatro atributos: *Beta*, *Sig.*, ordem e *R*; sendo que seus valores estão nas tabelas correspondentes no anexo B.

O *Beta* é o coeficiente padronizado da equação da regressão linear. Os valores deste coeficiente determinam a influência de cada variável independente na variável dependente. Ou seja, quanto maior for o valor de *Beta* que o critério assumir, maior será sua influência sobre a variável dependente. Para fins estatísticos, a *Sig.* (significância) dos valores de *Beta* adotados nos testes é de 0.05, ou melhor, margem de erro de 5%. Quanto ao *R*, este coeficiente mede a tendência de associação entre as variáveis, assumindo valores entre zero (nenhuma associação) e um (associação total). Neste projeto, os valores de *R* obedecem a

seguinte classificação: 0 a 0,33 (baixa associação); 0,34 a 0,75 (média associação) e 0,75 a 1 (alta associação). Com estas informações, pode-se analisar as tabelas a seguir.

Inicialmente, a Tabela 13 demonstra os resultados encontrados para os critérios de topicalidade intelectual.

Tabela 13 – Influência dos critérios de topicalidade intelectual nos espaços de relevância

Critérios	Espaço de relevância <i>Ex ante</i>					Espaço de relevância <i>Ex post</i>				
	1	2	3	4	5	1	2	3	4	5
1. Concordância entre os termos da consulta efetuada e o assunto apresentado pelo documento.	4°	5°	3*	1*	1*	1*	5°	4°	2*	4°
2. Informatividade do <i>abstract</i> para sua decisão de aprofundar-se no texto completo.	3°	3*	5	4*	4°	2°	1*	3*	5°	2*
3. Concordância entre as informações apresentadas no documento e sua necessidade de informação.	1°	4°	1*	2*	3°	3°	2*	1*	3*	1*
4. Profundidade/ foco do conteúdo no documento.	2°	1*	4°	5°	5°	4°	3°	2*	1*	5°
5. À atualidade do assunto apresentado.	3°	2*	2*	3*	2°	5°	4°	4°	4°	3°

* Significante ao nível de 0,05

OBS: Ver valores de R, *Beta* e significância no Anexo B: Tabela 19 (para *ex ante*) e Tabela 20 (para *ex post*)

Tendo uma visão geral desta Tabela, pode-se verificar a mudança de percepção do usuário em relação aos julgamentos de relevância sobre esta manifestação. Quanto à primeira tomada de medidas (*ex ante*), quando um documento era julgado como não relevante (valor 1) nenhum dos critérios acima apresentou significância estatística. Quando se passa para o espaço de quase não relevância (valor 2), o primeiro critério de influência diz respeito à profundidade do documento (critério 4) e, posteriormente, a atualidade do assunto (critério 5). Isso significa que para um objeto informacional ser julgado como quase não relevante basta ter tratamento superficial do assunto e/ ou o tema apresentado não ser atualizado.

Para um OI pertencer à região moderada de relevância, basta que exista uma concordância parcial entre as informações apresentadas no documento e a necessidade de informação, vindo em seguida os critérios referentes à atualidade e à concordância com os termos da consulta. Interessantemente, no momento *ex ante*, o fator de relevância preponderante para que um documento seja julgado como relevante (valor 4) e/ ou altamente relevante (valor 5) é a concordância entre os termos da consulta e o assunto apresentada pelo documento.

No que diz respeito aos dados *ex post*, pode ser verificado que na maioria dos julgamentos, a ordem dos fatores é radicalmente modificada em relação à medida *ex ante*. Nesta medida, para um documento ser julgado como não relevante bastam os termos da consulta não se associarem ao assunto apresentado (critério 1). Se a informatividade do abstract (critério 2) expressa algo um pouco diferente do que apresentado na consulta, acrescido de pouca concordância com a necessidade informacional, isto é o suficiente para o usuário julgar o documento como parcialmente não relevante (espaço de relevância= 2). Estes fatores estão mais de acordo com as estratégias de busca definidas pelos usuários.

Por outro lado, um documento é considerado moderadamente relevante (espaço de relevância= 3) se existe uma concordância parcial entre as informações do documento e a NI do usuário. Quanto às regiões mais altas de relevância, à profundidade do conteúdo (critério 4) é o aspecto principal para um OI ser considerado relevante e a concordância com a necessidade informacional (critério 3) é o fator essencial para um OI ser considerado altamente relevante (espaço de relevância= 5). Como aspecto principal, estes dados revelam um lado interessante de algumas pesquisas em SRI. Como visto na Tabela 13, no primeiro momento para um objeto ser julgado relevante e/ ou altamente relevante o critério principal de influência é a concordância dos termos da consulta e, como já discutido, normalmente este é o fator considerado principal para os projetos e avaliações de SRIs. No entanto, quando são mostrados os dados *ex post*, este critério não é o aspecto principal das regiões mais altas de relevância e sim, a concordância com a necessidade de informação.

Obviamente, esta diferenciação entre as classificações de relevância deve-se às formas de análise dos documentos. Mais uma vez, vale a pena ressaltar que o momento *ex ante* baseia-se na visão geral e superficial que o usuário tem do documento, diferentemente da visão *ex post* que é uma visão total e detalhada do OI e que retrata com fidelidade o que os usuários percebem de positivo e/ ou negativo nos objetos informacionais. Infelizmente, muitas das considerações dos projetos e avaliações de SRI levam em conta a primeira visão do usuário, quando assim o fazem. Desta forma, estes projetos baseiam-se em fatores de relevância incongruentes com a realidade. Portanto, a partir de então, será dada uma ênfase especial aos fatores de relevância *ex post* por estes se caracterizarem como julgamentos mais precisos do que os *ex ante*.

Outro aspecto interessante dos dados *ex post* é sua aparente estruturação lógica crescente. Ou seja, os julgamentos nos espaços de relevância vão mudando gradativamente, levando em consideração a influência do fator principal da região anterior. Isto pode ser melhor visualizado na Figura 11 a seguir.

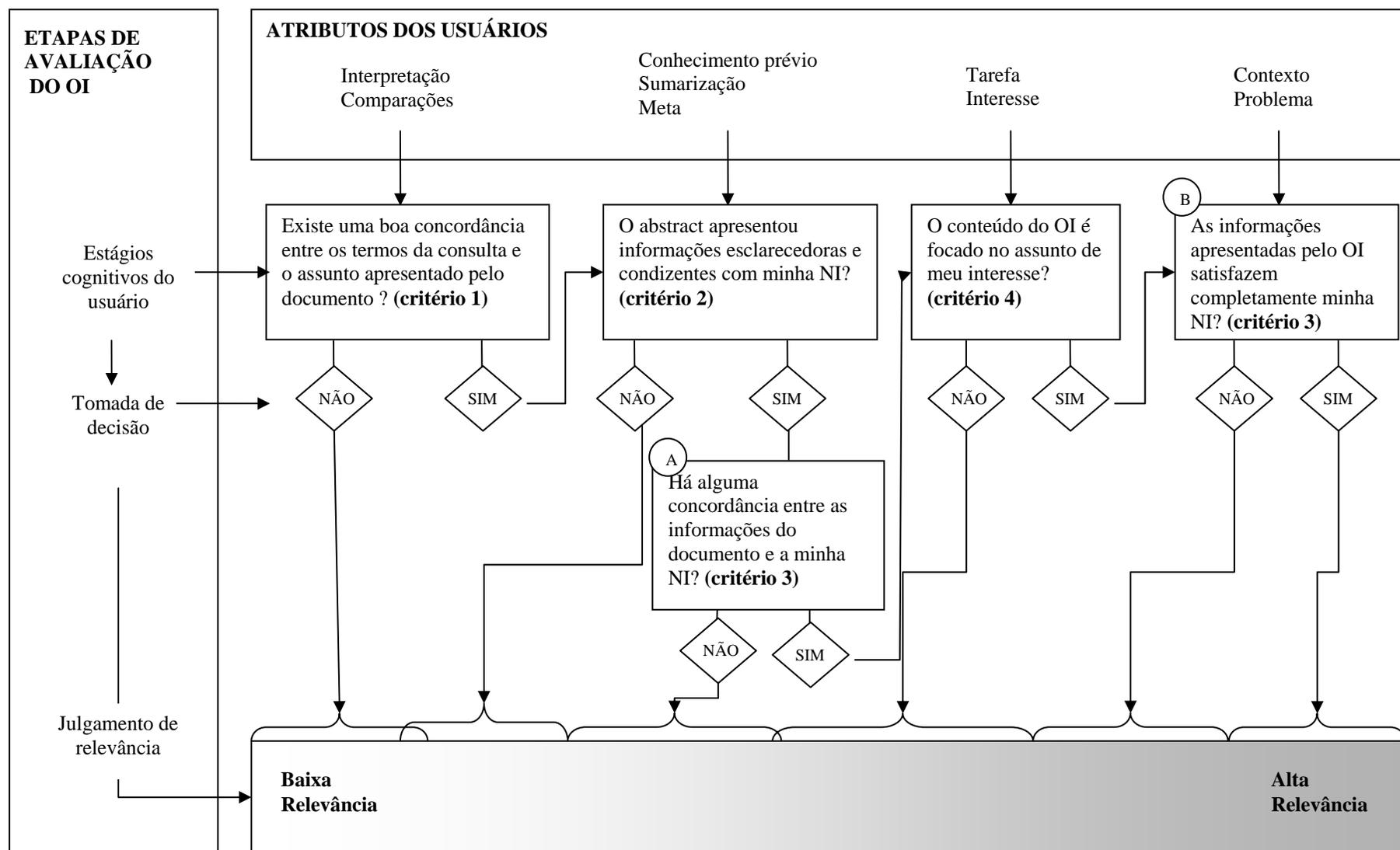


Figura 10 - Fatores de relevância para a topicalidade intelectual

No lado esquerdo da Figura 13 está o processo de avaliação dos OIs. Este processo inicia-se com os estágios cognitivos dos usuários (perguntas internas sobre os critérios de relevância). Como consequência, existe uma tomada de decisão em relação ao critério de relevância e, por fim, existe o julgamento do OI em algum espaço de relevância. Na parte superior, estão demonstrados os atributos que influenciam os usuários durante o processo de RI e julgamento de relevância. Normalmente envolve aspectos situacionais, cognitivos e psicológicos.

Na parte inferior, estão os espaços ou regiões de relevância de acordo com a escala utilizada no instrumento de pesquisa (baixa relevância até alta relevância). Por fim, a parte central e mais importante, diz respeito ao processo gradativo e acumulativo do julgamento, onde cada decisão influencia as decisões posteriores. Em linhas gerais, o processo flui da seguinte forma. Inicialmente, o usuário se questiona se existe alguma concordância entre os termos da consulta (critério 1) e o assunto apresentado. Se houver pouca ou nenhuma concordância, isto torna-se o suficiente para o usuário avaliar o objeto em algum grau do espaço de baixa relevância. Caso haja alguma concordância, o usuário questiona-se sobre o próximo critério de importância para ele que, no caso desta manifestação foi o critério 2. Esta sequência lógica vai sendo utilizada pelo usuário até ele obter uma decisão final sobre o objeto informacional.

Também na mesma Figura, pode-se perceber que o critério 3 aparece duas vezes, representados pelas letras A e B. Isto advém dos dados da Tabela 13, em que este critério é o primeiro em importância em duas regiões, e também da observação e citação dos julgamentos efetuados pelos respondentes. No caso, os usuários descreviam que analisavam alguns atributos dos OIs mais de uma vez para se chegar a uma decisão final. Por exemplo, se o usuário percebesse que existia alguma concordância entre as informações contidas no OI e a necessidade de informação, isto já era o suficiente para ele passar para o próximo estágio e, assim, analisar se o conteúdo era focado no assunto de interesse (critério 4). Depois de verificar este aspecto, o usuário mais uma vez recorria ao critério 3 para verificar a concordância total com sua necessidade de informação e, por fim, obter uma decisão final sobre o objeto informacional.

Em suma, com esta Figura foi possível perceber mais claramente os fatores de relevância para a manifestação denominada topicalidade intelectual.

No que diz respeito à manifestação referente à estrutura/ formato do OI, seus resultados podem ser vistos na Tabela abaixo.

Tabela 14 – Influência dos critérios de estrutura nos espaços de relevância

Critérios	Espaço de relevância <i>Ex ante</i>					Espaço de relevância <i>Ex post</i>				
	1	2	3	4	5	1	2	3	4	5
7. Clareza apresentada no documento.	2°	2*	3*	1*	2*	3°	3*	3*	1*	2*
8. Credibilidade (informações corretas e válidas) provida pelo documento.	1°	3*	2*	3*	1*	1°	2*	1*	3*	3*
9. Efetividade do método ou procedimento utilizado pelo autor.	3°	1*	1*	2*	3*	2°	1*	2*	2*	1*
10. Facilidade de obtenção do documento.	4°	4°	4°	4*	4*	4°	4°	4°	4°	4*

* Significante ao nível de 0,05

OBS: Ver valores de R, *Beta* e significância no Anexo B: Tabela 21 (para *ex ante*) e Tabela 22 (para *ex post*)

Como pode ser observado, de uma forma geral tanto na medida *ex ante* como na medida *ex post*, a efetividade do método ou procedimento (variável 9) é o critério que mais influência possui nos espaços de relevância. Quanto à facilidade de obtenção do documento (variável 10), este critério não apresentou muita importância para a decisão do usuário, já que isto era considerado como fator do processo de busca, ou seja, vinha antes do julgamento do objeto informacional. E, de uma forma geral, nenhum critério apresentou significância estatística em relação à consideração do documento como não relevante (espaço de relevância= 1).

Observando os dados *ex post*, quando o documento é avaliado como parcialmente não relevante (espaço= 2), o primeiro critério a considerar é a efetividade do método apresentado no documento como um fator que não satisfaz o usuário, vindo em seguida uma baixa credibilidade das informações contidas no documento (critério 8) e, por fim, falta de maior clareza de conteúdo OI (critério 7). Quando o objeto é considerado de relevância moderada (valor 3), o fator preponderante a ser averiguado é a credibilidade das informações providas pelo objeto (critério 2). Se estes critérios forem satisfeitos positivamente, a próxima etapa é verificar a credibilidade das informações. Caso não seja o suficiente, isto torna-se um fator principal para o usuário atribuir relevância moderada ao objeto.

Por fim, têm-se as regiões superiores de relevância. Interessantemente, há uma troca de posições entre o primeiro e segundo lugares em ambas regiões. Considerando a região de alta relevância, a efetividade do método ou procedimento utilizado pelo autor do OI é o critério de relevância que mais importância tem para os usuários, seguido de perto pela

clareza apresentada no documento. Mais uma vez, da conjuntura entre dos dados *ex post*, definições das estratégias e relatos dos respondentes, pode-se levantar um procedimento de julgamento de relevância, exemplificado na Figura abaixo.

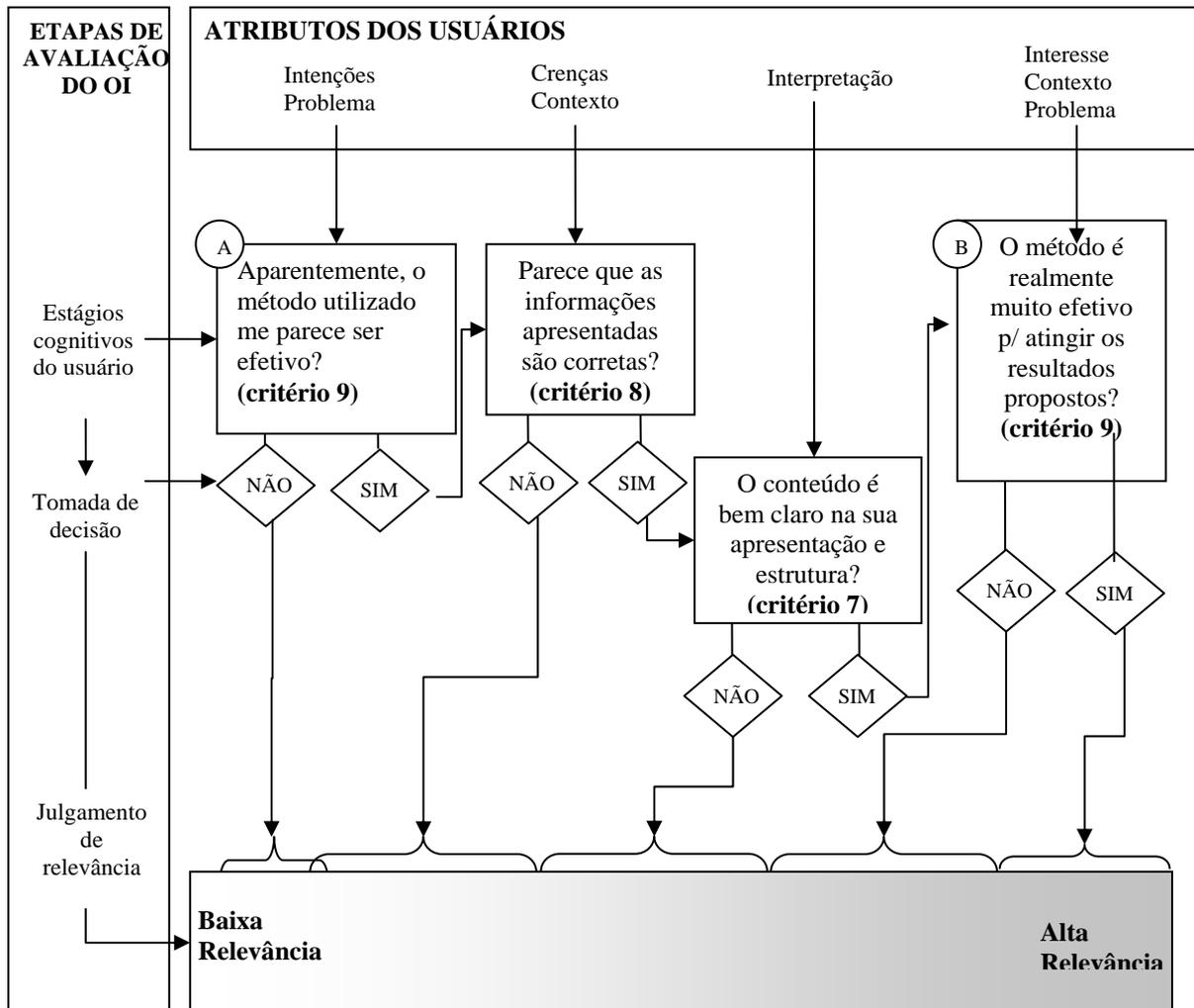


Figura 11 – Fatores de relevância para estrutura/ formato

As partes pertencentes a esta Figura seguem o mesmo padrão da Figura 1, explicado anteriormente. Obviamente, as modificações encontradas referem-se aos critérios de relevância e aos atributos dos usuários. Vale observar a recorrência ao critério 9, referente a efetividade do método, sendo isto representado na Figura 12 com as letras A e B. Quanto à letra A, o usuário utiliza a efetividade do método inicialmente como um ponto de corte, ou seja, caso este atributo do OI não aparente ser efetivo, o usuário já atribui um baixo valor de relevância. Senão, ele passa para os próximos estágios até chegar na letra B. Neste caso, o usuário voltou a analisar mais uma vez a efetividade do método para ter uma decisão final sobre o OI.

Em relação à manifestação cognitiva ou pertinência informacional, a Tabela abaixo mostra os resultados da regressão linear múltipla.

Tabela 15 - Influência dos critérios cognitivos nos espaços de relevância

Critérios	Espaços de relevância <i>Ex ante</i>					Espaços de relevância <i>Ex post</i>				
	1	2	3	4	5	1	2	3	4	5
12. Aumento de conhecimento acerca de sua necessidade de informação.	2°	5°	3*	3*	4°	5°	4°	3*	4°	1*
13. Obtenção de novas informações sobre o assunto pesquisado.	1*	1*	4°	2*	2*	3°	1*	2*	5°	2*
14. Ampliação de novas perspectivas de conhecimento.	4°	3°	1*	5*	3°	4°	3*	1*	1*	3*
15. Conhecimento adquirido sobre o assunto pesquisado.	5°	4°	5°	4*	1*	2*	5°	4°	2*	5°
16. Reforço ou suporte ao seu ponto de vista sobre o assunto pesquisado.	3°	2°	2*	1*	5°	1*	2*	5°	3*	4°

* Significante ao nível de 0,05

OBS: Ver valores de R, *Beta* e significância no Anexo B: Tabela 23 (para *ex ante*) e Tabela 24 (para *ex post*)

Analisando os dados de uma forma geral, a obtenção de novas informações sobre o assunto pesquisado (critério 13), é o critério mais lembrado pelos respondentes quando se pensa nos espaços inferiores de relevância. Para os demais critérios, existe uma modificação substancial da ordem dos fatores nas duas medições. Observando-se os valores referentes às medidas *ex post*, pode-se observar que o aspecto principal para o OI ser julgado como não relevante (valor 1) é a falta de reforço ou suporte ao ponto de vista que o usuário possui sobre o assunto (critério 16).

Quando um objeto informacional é julgado como parcialmente não relevante (valor 2), o primeiro fator de influência é o critério 13. Ou seja, provavelmente o conteúdo deste objeto apresentou pouca ou nenhuma informação nova sobre o assunto. Em seguida, vem o critério 16 como mais um ponto de corte e, por fim, a pouca ampliação de novas perspectivas de conhecimento sobre o tema abordado (critério 14).

Este último fator torna-se o aspecto primordial quando um documento é tido como moderadamente relevante. Ou seja, pode existir alguma abertura sobre novos enfoques interessante sobre o assunto, mas não o suficiente ou esperado pelo usuário. O segundo fator de influência diz respeito a uma quantidade razoável de novas informações apresentadas pelo OI e, por último, o aumento um pouco significativo de conhecimento acerca da necessidade de informação do usuário (critério 12).

Interessantemente, o mesmo critério de importância da região de relevância moderada (valor 3) manteve-se como critério que mais influência tem para um OI ser julgado como relevante (valor 4). A segunda importância para um objeto ser julgado relevante diz respeito ao conhecimento adquirido sobre o assunto pesquisado e, em conjunto, o reforço ou suporte ao ponto de vista do usuário sobre o assunto abordado.

Quanto ao espaço de alta relevância, o critério mais importante é o aumento de conhecimento acerca da necessidade de informação do usuário (critério 12), vindo em seguida a obtenção de novas informações sobre o assunto pesquisado e, por fim, a ampliação de novas perspectivas. Como podem ser visualizados na Tabela 15, os dados mantiveram um alinhamento nas duas medições no que diz respeito aos primeiros critérios de influência sobre a região de quase não relevância e a região moderada. No entanto, quando se trata das regiões superiores de relevância, os julgamentos nas duas medidas praticamente assumem ordens quase invertidas. Basta verificar que o conhecimento adquirido é o fator primordial na análise *ex ante* (valor 5) e, passa a ser o último fator de importância na análise *ex post* para se julgar um documento como altamente relevante.

Na verdade, isto é influenciado principalmente pelas características intrínsecas da manifestação cognitiva e seus critérios. Certamente, é muito mais difícil capturar estes aspectos com uma visão superficial, como é a *ex ante*. Seguindo a tendência das manifestações anteriores, os julgamentos parecem ter um aspecto crescente de comportamento. A Figura a seguir mostra este comportamento baseado nos julgamentos *ex post*, por traduzirem com maior fidelidade a percepção cognitiva.

Mais uma vez, a estrutura da Figura a seguir é semelhante às das demais manifestações. As diferenças dizem respeito aos atributos dos usuários e aos critérios de relevância. Pode-se também observar que todos os critérios foram envolvidos no processo de julgamento e, em especial, o critério 15 (conhecimento adquirido sobre o assunto pesquisado) por este ser uma extensão do critério 14 na região de OIs considerados relevantes (valor 4).

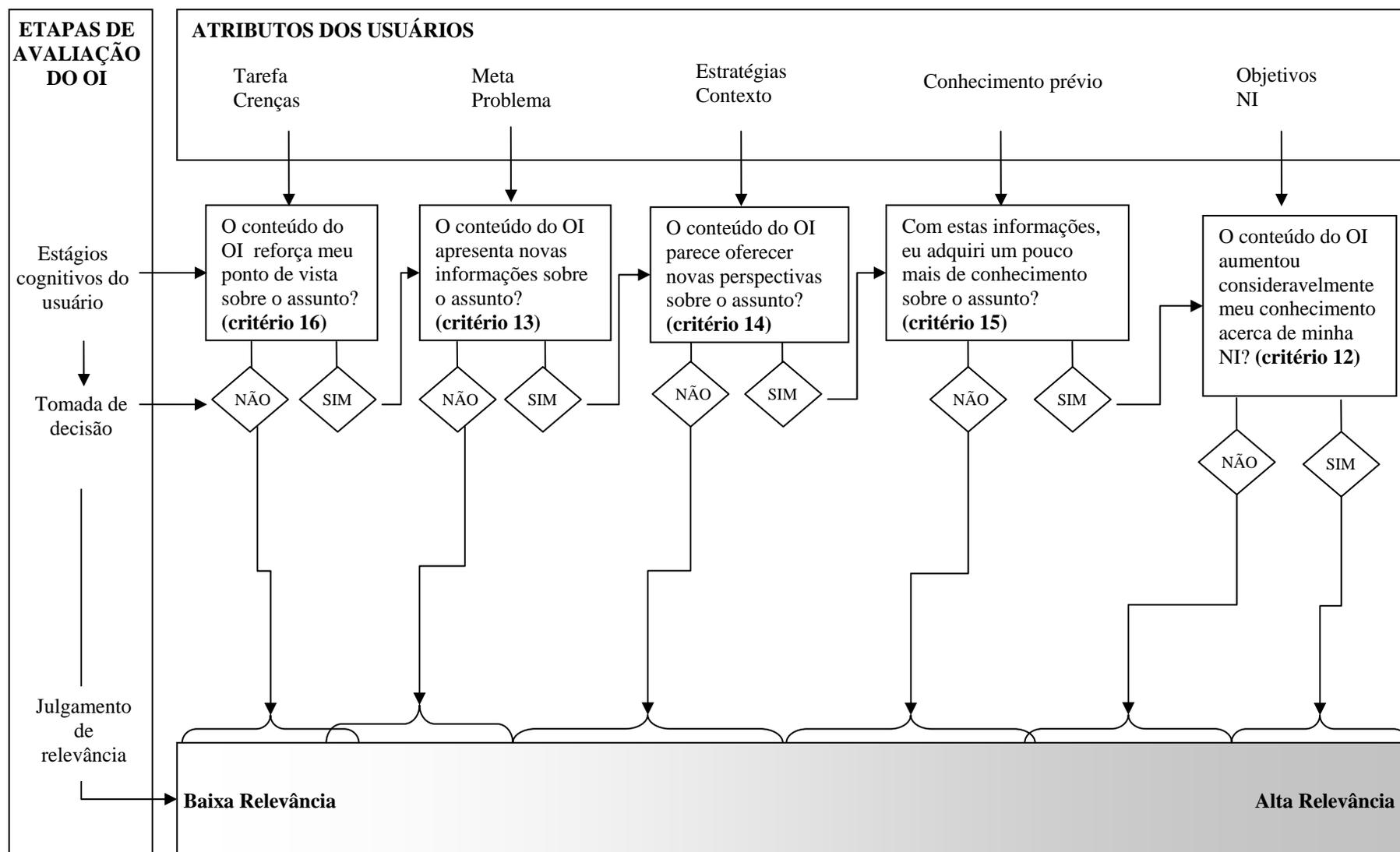


Figura 12 - Fatores de relevância para a manifestação cognitiva

Em relação à manifestação denominada utilidade, a Tabela 16 mostra os dados obtidos após a regressão linear.

Tabela 16 - Influência dos critérios de utilidade nos espaços de relevância

Critérios	Espaço de relevância <i>Ex ante</i>					Espaço de relevância <i>Ex post</i>				
	1	2	3	4	5	1	2	3	4	5
18. Utilidade para sua necessidade de informação.	-	3*	3*	2*	2°	-	2*	1*	1*	5°
19. Apresentação de novidades úteis para sua necessidade de informação.	-	2*	4°	4°	1°	-	4°	4°	4*	4°
20. Utilidade na resolução do problema informacional que você tem em mãos.	-	1*	2°	1*	3°	-	5°	5°	2*	1*
21. Utilidade para tomar uma decisão sobre a necessidade de informação necessária.	-	5°	1*	5°	4°	1°	1°	2°	5°	2°
22. Utilidade para reduzir alguma incerteza quanto à sua necessidade de informação.	-	4°	5°	3°	5°	-	3°	3°	3*	3°

* Significante ao nível de 0,05

OBS: Ver valores de R, *Beta* e significância no Anexo B: Tabela 25 (para *ex ante*) e Tabela 26 (para *ex post*)

Como pode ser visto, para o espaço baixa relevância (valor 1) os dados não apresentaram significância estatística nas duas tomadas de medições. Na análise *ex ante*, os julgamentos em relação a este espaço não apresentaram variabilidade de julgamento. Para análise *ex post*, até mesmo o critério 21 não apresentou significância aceitável. Observando os dados que tiveram significância estatística nas duas medições, o critério 20 é o fator de importância mais lembrado na análise *ex ante* e o critério 18 é o mais influente na análise *ex post*. No entanto, nenhum dos dois é referenciado para o nível de mais alta relevância.

Dando ênfase aos dados *ex post*, o primeiro critério que apresentou significância foi a utilidade para a necessidade de informação (critério 18). Este mesmo critério passa a ser o fator principal na região moderada de relevância. Ou melhor, se a utilidade para a NI do usuário é atendida apenas em parte, isto é o suficiente para o objeto ser considerado moderadamente relevante.

No que diz respeito ao espaço dos documentos considerados relevantes (valor 4), o mesmo critério acima citado permanece como fator principal, significando que a utilidade do documento para a NI do usuário foi atendida. Em segundo lugar de influência, vem a utilidade na resolução do problema informacional que o usuário possui em mãos (critério 20).

Se comparada às outras manifestações, a utilidade foi a que obteve maiores mudanças de julgamentos de relevância entre os dois momentos, sendo que os critérios 18 e 20 foram os que demonstraram influenciar essencialmente nas decisões dos usuários, conforme os dados da Tabela anterior. Obviamente, esta é uma das manifestações mais importantes de relevância e depende fortemente do contexto em que o usuário está inserido, seus objetivos, metas e intenções. Certamente, esta variabilidade de percepção advém destes atributos e da característica das duas etapas. Por isto, a importância dos resultados *ex post* serem levados em consideração para a avaliação dos SRI.

Esta manifestação também apresentou uma característica própria de julgamento. Na Figura 14 a seguir, pode ser visto os atributos dos usuários que influenciam o processo de julgamento de relevância e os critérios envolvidos.

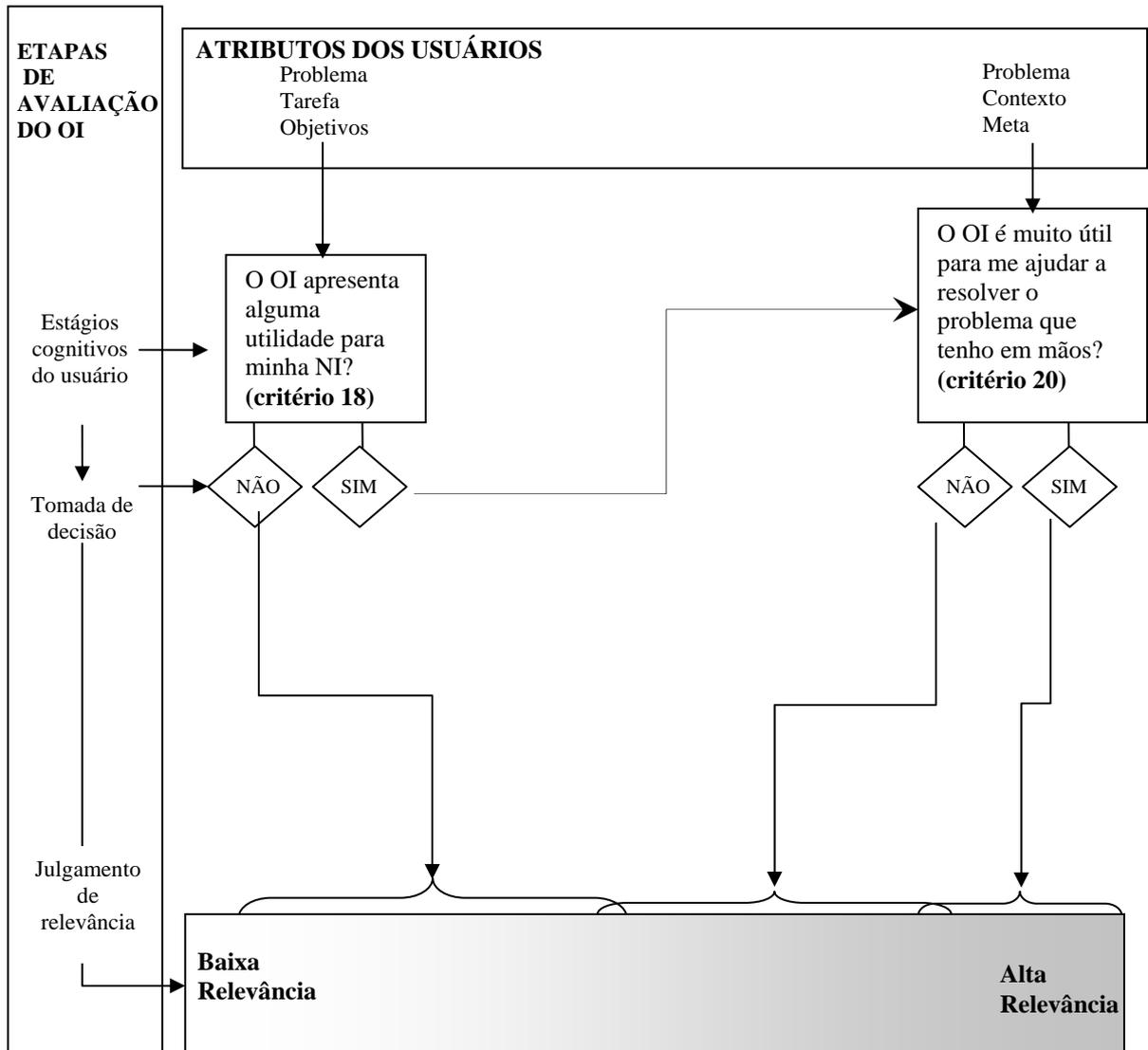


Figura 13 - Fatores de relevância para a utilidade

Finalmente, a Tabela 17 mostra os resultados da última manifestação estudada, a motivacional.

Tabela 17 - Influência dos critérios motivacionais nos espaços de relevância

Critérios	Espaços de relevância <i>Ex ante</i>					Espaços de relevância <i>Ex post</i>				
	1	2	3	4	5	1	2	3	4	5
24. Estímulo que o conteúdo do documento lhe ofereceu a aprofundar-se mais no assunto.	2	1*	3	1*	2*	3	1*	1*	3*	1*
25. Inspiração a pesquisar outros objetivos além de sua necessidade de informação.	3	4	1*	4*	3	1*	4	3*	4*	4
26. Motivação a procurar pesquisas feitas pelo mesmo autor.	-	2	4	3*	4	-	3*	4	2*	2*
27. Motivação a procurar os artigos citados pelo autor.	1	3	2	2*	1*	2	2*	2*	1*	3*

* Significante ao nível de 0,05

OBS: Ver valores de R, *Beta* e significância no Anexo B: Tabela 27 (para *ex ante*) e Tabela 28 (para *ex post*)

Como pode ser identificado acima, o estímulo que o conteúdo do documento ofereceu para o usuário aprofundar-se mais no assunto (critério 24) foi o fator que mais influenciou nas regiões de relevância da manifestação motivacional. Observando os dados *ex ante*, este critério somente não influenciou consideravelmente a região moderada.

Quando analisados os dados *ex post*, este mesmo critério tornou-se fator essencial em três regiões e, principalmente, na região de mais alta relevância. O segundo critério da manifestação motivacional que mais influenciou no julgamento de relevância foi a motivação a procurar os artigos citados pelo autor (critério 27), o que entra em concordância com uma das estratégias de buscas do usuário. Este mesmo procedimento é observado nos dados *ex ante*, só que neste caso, o espaço em que isto se verificava é o de alta relevância.

Seguindo a tendência das manifestações anteriores, a motivacional também possui uma estrutura lógica de decisão para os fatores de relevância e isto pode ser visto na Figura 15.

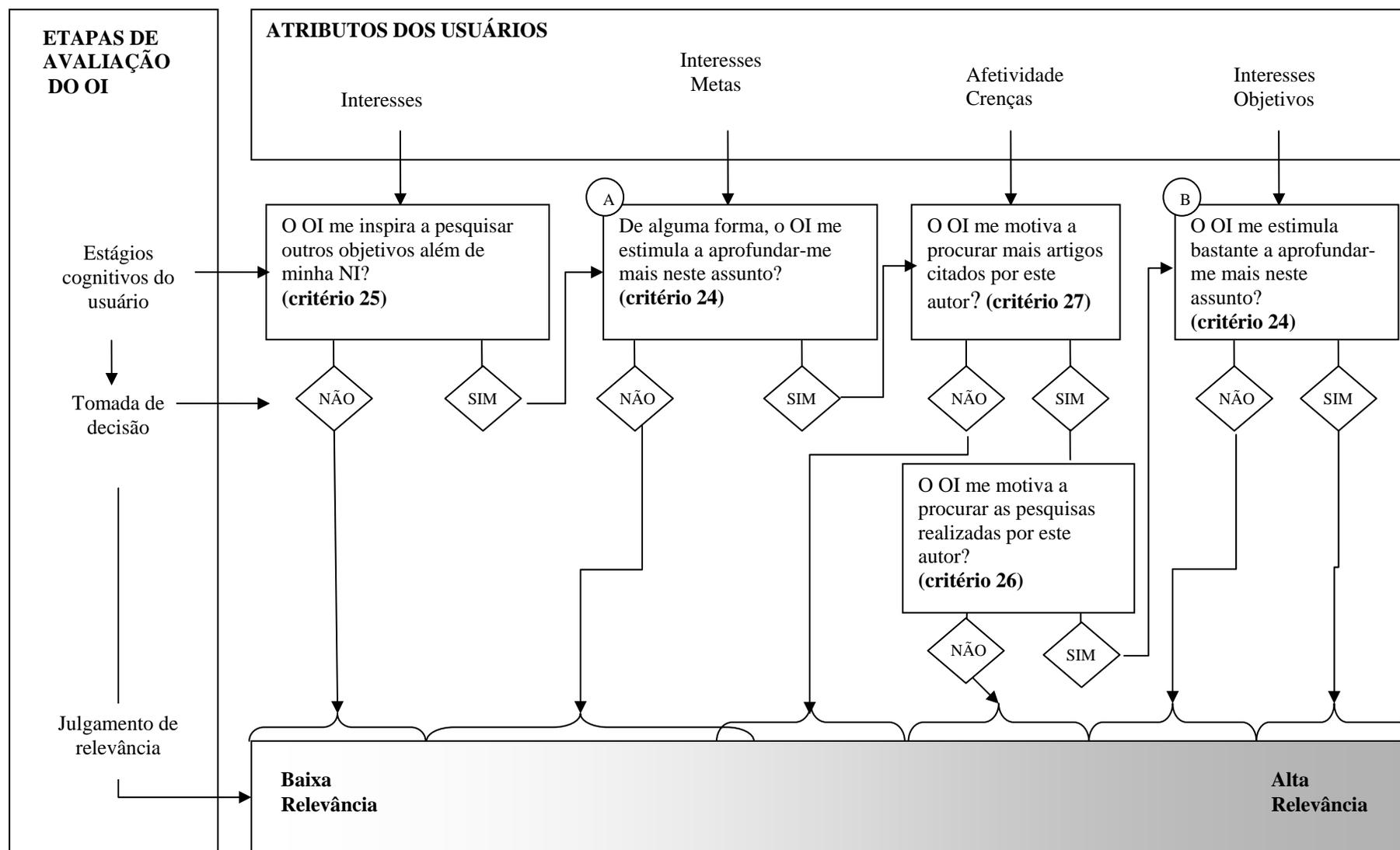


Figura 14 - Fatores de relevância para a manifestação motivacional

Na Figura 15 também pode ser visto a recorrência ao critério 24 (estímulo a aprofundar-se mais no assunto) representada pelas letras A e B. Aqui, o usuário utilizou este critério mais de uma vez para decidir sobre a relevância do OI e isto pode ser visto na importância deste critério na Tabela 16 (dados *ex post*). Inicialmente, o usuário se questiona se o OI transmite informações que o estimula a aprofundar-se mais no assunto. Se isto acontecer ele prossegue sua análise do OI (Letra A). Finalmente, depois de ter passado pelos outros estágios, o usuário recorre ao critério 24 para verificar se o OI lhe estimula bastante, ou seja, já admite o aspecto positivo do critério. Caso isto aconteça (na letra B), o OI é considerado em algum grau de alta relevância. Senão, se o OI o estimula a aprofundar-se no assunto, mas não extremamente, o OI é julgado no espaço considerado relevante.

Enfim, as Tabelas antecedentes (13 a 17) mostraram os fatores de relevância para o julgamento de um objeto informacional. Estes fatores são importantes para a especificação de requisitos de SRIs porque demonstram os atributos que os usuários utilizam durante o processo de recuperação de informação. São também importantes para avaliação do desempenho destes sistemas, já que como visto nestas Figuras, os usuários empregam vários recursos de julgamento muito mais condizentes com a sua realidade do que a relevância sistêmica.

Porém, para completar a conjuntura dos fatores de relevância é necessário saber a ordem de importância que as manifestações possuem para o julgamento de relevância. Este resultado é mostrado na Tabela 18 a seguir.

Tabela 18 - Percepção global da relevância de acordo com a manifestação

Geral	<i>Ex ante</i>			<i>Ex post</i>		
	Beta	Sig.	Ordem	Beta	Sig.	Ordem
Topicalidade intelectual	0,170	0,000	3*	0,181	0,000	4*
Estrutura/ formato	0,010	0,736	5°	0,055	0,065	5°
Cognitivo/ pertinência	0,116	0,023	4*	0,222	0,000	2*
Utilidade	0,472	0,000	1*	0,394	0,000	1*
Motivacional	0,259	0,000	2*	0,189	0,000	3*

* Significante ao nível de 0,05

$R_{ex\ ante} = 0,914$ $R_{ex\ post} = 0,935$

Como podem ser visualizadas, as únicas manifestações que mantiveram suas posições de influência no julgamento de relevância do documento foram a utilidade e a estrutura/ formato. Ou seja, a utilidade é a manifestação lembrada em primeira instância quando o usuário necessita julgar a relevância de um objeto informacional. Já a manifestação estrutura/ formato do OI não apresentou valores aceitáveis significância para análise em ambas medições.

Quanto aos dados *ex ante*, é possível identificar a influência da estratégia de busca baseada na afetividade com o autor, já que a manifestação motivacional assume a segunda ordem de importância nesta primeira medição. Já a topicalidade intelectual assume o terceiro posto, certamente devido também ao processo de busca no qual os usuários utilizavam mais o julgamento de concordância informacional para descartar um objeto do que para julgá-lo como relevante. E, surpreendentemente, a manifestação cognitiva ou pertinência informacional vem na penúltima posição, provavelmente devido à análise superficial da primeira medida. Na verdade, o julgamento *ex ante* é um processo importante, porém limitado em termos da representatividade dos fatores de relevância.

Quando se visualizam os dados *ex post*, há uma mudança significativa na ordem de influência da manifestação. A manifestação referente aos aspectos cognitivos ou pertinência informacional passou de quarto para o segundo maior fator de julgamento de relevância. Já as manifestações motivacional e a topicalidade intelectual caíram uma posição em relação à medida *ex ante*. Em ambas as medidas, pode ser visto que a topicalidade (mesmo a

intelectual), não é um fator essencial no processo de RI; pior ainda se for considerado a topicalidade ou relevância sistêmica.

Como já elucidado anteriormente, os dados *ex post* retratam com mais precisão o julgamento do usuário. Portanto, será dada uma ênfase maior à discussão destes dados da Tabela 18. A Figura 16 a seguir, ajuda no entendimento desta discussão.

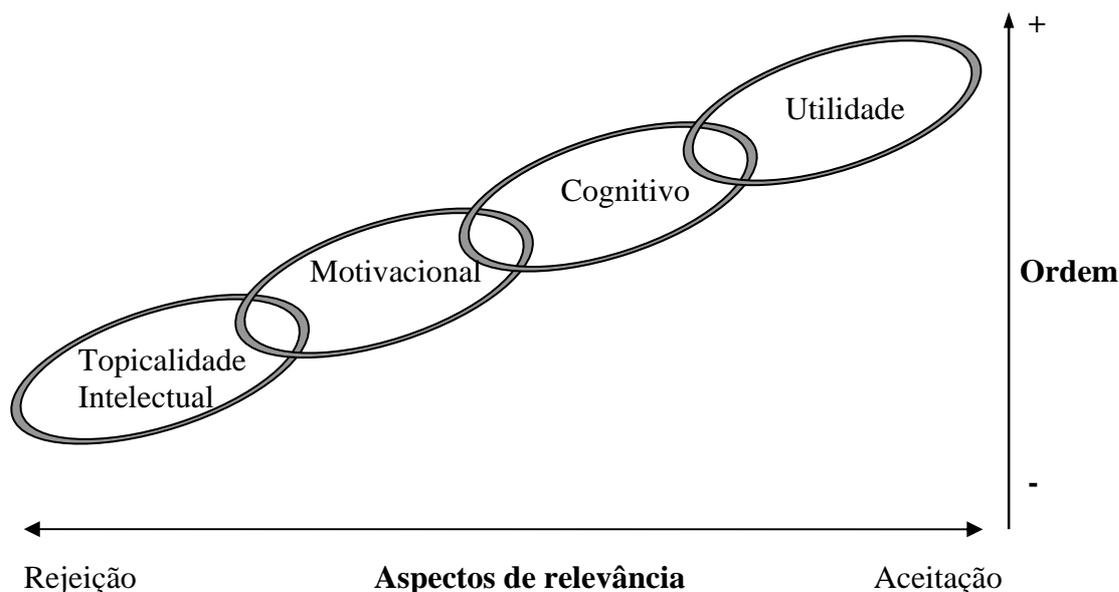


Figura 15 - Visão geral das manifestações de relevância

Como visto na Figura acima, a manifestação denominada estrutura/ formato não está presente por não apresentar significância estatística. As demais manifestações estão organizadas em ordem crescente de importância, obedecendo aos dados *ex post* da Tabela 18. As manifestações de topicalidade intelectual e motivacional estão no posicionamento que refere-se ao julgamento dos aspectos negativos da relevância (espaços ou regiões inferiores). Isto porque boa parte dos critérios destas manifestações servia como ponto de corte durante a seleção dos objetos informacionais como, por exemplo, a verificação da concordância entre os termos da consulta e o assunto do documento, e a afinidade com o autor.

Quanto aos aspectos positivos de relevância, devem-se observar as manifestações de utilidade e cognitiva. Ou seja, quando se pensa em altos níveis de relevância, o primeiro aspecto que os usuários desejam que o OI possua são os critérios de utilidade e os cognitivos. Obviamente, estas manifestações são fatores complexos de se mensurar. Disto advém a importância das considerações dos usuários durante a avaliação dos sistemas de recuperação de informação.

Em suma, o projetista de SRI deve ter em mente a importância destas manifestações e analisar os fatores de relevância para cada uma delas. Por fim, de posse dos resultados referentes à definição da estratégia, formulação da consulta e os fatores de relevância, a próxima etapa é elucidar as conclusões deste trabalho e suas implicações, assuntos do próximo capítulo.

7 CONSIDERAÇÕES FINAIS

Este capítulo apresenta as conclusões obtidas nesta pesquisa, as implicações, limitações e sugestões para futuras pesquisas.

7.1 CONCLUSÕES

A busca por sistemas de recuperação de informação cada vez mais eficientes é um desafio que envolve diversas áreas do conhecimento tais como, sistemas de informação, ciência da informação, psicologia, entre outras. Muito foi e tem sido feito pelos aspectos computacionais destes sistemas como, por exemplo, desenvolvimento de algoritmos de busca e interfaces amigáveis. No entanto, os aspectos considerados pelos usuários ainda não foram agregados coerentemente aos projetos de SRI, muito em consequência da carência de estudos com esta abordagem.

Então, notando a possibilidade de se estudar os aspectos de relevância pertencentes aos usuários, o objetivo principal deste estudo focou essencialmente no processo de recuperação de informação com o intuito de se explorar os fatores de relevância que os usuários utilizam para julgar os objetos informacionais. Para se obter estes fatores, esta pesquisa foi desenvolvida em etapas que permitiram o entendimento das características que os usuários utilizam durante o processo de recuperação de informação.

A primeira etapa, que também era um dos objetivos específicos, mostrou as definições de estratégia de busca que os respondentes utilizavam antes de efetuar a consulta, sendo que quatro grandes táticas foram levantadas e analisadas de acordo com o perfil do respondente. Destas, a abrangência com o tema em pesquisa e a baseada na afetividade com o autor eram as mais utilizadas. Em seguida, vinham as estratégias baseadas no reconhecimento da fonte de publicação e nas citações dos autores de afinidade. Foi percebido que o ano (ou idade) do objeto informacional era um fator delimitador na estratégia de busca e isto era amplamente utilizado pelos respondentes. Esta etapa mostra-se muito importante para o processo de exploração dos fatores porque alguns critérios de relevância já são empregados desde esta etapa, tais como, afinidade com autor e tempo do OI.

A segunda etapa, também um objetivo específico desta pesquisa, é decorrente da anterior e tinha como tarefa, observar as táticas de formulação de consultas utilizadas pelos respondentes. Foram identificados três que, de acordo com sua lógica, foram denominadas como *top-down*, *middle* e *bottom-up*. Esta etapa era importante para seleção dos OIs que interessavam aos usuários. Verificou-se então, que os usuários utilizavam um conjunto de passos bem definidos para analisar o OI. Este conjunto de passos revelou que a concordância entre os termos da consulta e o assunto pesquisado era utilizado mais para eliminar um documento do que para considerá-lo como relevante. Este aspecto tem profundo impacto nos projetos de SRI já que a concordância tópica é utilizada como parâmetro de sucesso de relevância, ou seja, o inverso do analisado nesta pesquisa.

Por fim, foi explorada a etapa mais importante, a análise dos objetos informacionais. Esta etapa ofereceu para esta pesquisa um dos objetivos específicos e o objetivo geral. O objetivo específico teve como resultado o levantamento de cinco novos critérios que não foram incluídos no instrumento de pesquisa. O critério referente a artigos de mercados emergentes, pode ser um atributo que deva ser levado em consideração para os mecanismos de busca do SRI possibilitando o usuário escolher a origem geográfica do estudo aplicado no artigo.

O objetivo geral (caracterização dos fatores de relevância) mostrou aspectos interessantes e também inusitados. Ficou evidenciado que os dados *ex post* possuem uma modificação de percepção acentuada em relação aos dados *ex ante*, sendo que os *ex post* revelam julgamentos de relevância mais confiáveis. Partindo deste princípio, observou-se que a utilidade é o aspecto mais importante quando se pensa em julgar um OI, seguida da manifestação cognitiva, depois a motivacional e, por fim, a topicalidade intelectual. Das observações, entrevistas, relatos dos usuários e análise estatística chega-se à conclusão que as manifestações de topicalidade intelectual e a motivacional servem mais como parâmetro para se ter a decisão de rejeitar ou não o artigo. Assim, quando os projetistas forem recorrer aos fatores de relevância destas duas manifestações, eles devem levar em consideração essencialmente os critérios que possuem maior influência nas regiões inferiores de relevância, já que eles acusam qual foi a falha de desempenho do SRI em recuperar tal objeto informacional.

Inversamente, os fatores de relevância da utilidade e da manifestação cognitiva devem ser levados em conta como atributos de sucesso do objeto informacional. Ou melhor, os fatores dos espaços superiores de relevância destas manifestações ditam a avaliação do objeto informacional como relevante ou altamente relevante.

Em suma, os fatores de relevância mostraram que os usuários utilizam um processo de julgamento crescente de relevância e, que as classificações são muito mais acuradas do que a dicotômica (relevante/ não relevante) utilizada pela maioria dos SRIs. Desta forma, esta pesquisa possui algumas implicações que são descritas a seguir.

7.2 IMPLICAÇÕES DA PESQUISA

7.2.1 Estudos sobre Relevância

Como abordado no capítulo referente à contextualização da pesquisa, os estudos de relevância estão progredindo em duas áreas desconectadas. Uma área foca no levantamento de critérios e a outra no estudo teórico dos espaços de relevância. Devido à esta conjuntura, esta pesquisa uniu estes enfoques (baseando-se nas manifestações de relevância) e observou que a correlação entre os critérios e os espaços de relevância, ou melhor, fatores de relevância, revelam um interessante comportamento decisório dos usuários durante o julgamento dos OIs. Isto é de extrema significância para os estudos de relevância porque o critério em si revela apenas um atributo esperado pelo usuário, mas quando analisado conjuntamente com o julgamento em espaços de relevância, os critérios revelam o comportamento dos usuários durante o processo de RI.

Outro aspecto interessante para os estudos de relevância diz respeito à percepção dos dados *ex ante* e *ex post*. Como discutido anteriormente, a análise *ex ante* é muito superficial e, portanto, condiz pouco com a realidade de julgamento do usuário. Isto é evidenciado na mudança de percepção dos fatores de relevância (Tabelas 13 a 17) e na importância das manifestações (Tabela 18). Infelizmente, os estudos de relevância são focados prioritariamente nos dados *ex ante*, recolhidos de uma única interação do usuário com o sistema. E, por consequência, estes resultados são levados em consideração para a definição dos requisitos de um SRI, sendo dada ênfase aos aspectos da topicalidade.

Outro aspecto analisado neste estudo refere-se ao levantamento de critérios de relevância. Os resultados mostraram que os critérios de relevância já podem ser identificados antes dos usuários interagirem com o sistema. Ou seja, os usuários não revelam seus critérios apenas quando interagem com o objeto informacional e sim, durante todo o processo de recuperação de informação. Isto pode abrir uma possibilidade interessante para se levantar e estudar novos critérios de relevância.

7.2.2 Projetos de Sistemas de RI

Os fatores de relevância encontrados nesta pesquisa demonstraram duas grandes limitações dos projetos de SRI. A primeira delas diz respeito ao baixo desempenho dos sistemas de SRI. Como discutido ao longo desta pesquisa, a classificação que os projetistas destes sistema adotam é a dicotômica (relevante ou não relevante) e, portanto, os modelos computacionais são desenvolvidos com base neste aspecto, onde a probabilidade de ocorrência dos termos da consulta é simplesmente uma função que leva em consideração esta classificação extremista da relevância.

Mas, como visto nesta pesquisa, os usuários entendem e adotam perfeitamente a classificação em mais espaços de relevância, levando em conta que a ocorrência dos critérios parciais de relevância influenciam no julgamento das regiões posteriores. Assim, uma sugestão para o desenvolvimento das técnicas e fórmulas probabilísticas que dão suporte aos algoritmos computacionais de RI, seria a inclusão de variáveis que representassem os espaços parciais de relevância. Isto poderia acarretar num processo de busca mais refinado, melhorando a etapa de ordenação dos OIS, já que até então, os SRIs agrupam as regiões parciais na região de baixa relevância e, assim, comprometem todo o processo de busca dos OIs das regiões superiores de relevância.

A segunda limitação diz respeito à especificação dos requisitos dos SRIs. Atualmente, sabe-se que os requisitos que são levados em conta referem-se à associação de termos da consulta. No entanto, os fatores de relevância explorados nesta pesquisa demonstraram que as listas de requisitos ou critérios que os usuários empregam é muito mais extensa do que uma simples associação de termos. Portanto, as fases dos projetos que visam adicionar os requisitos e avaliar os protótipos do SRI devem levar mais em consideração as classes de avaliações que incluem o uso e o usuário e os aspectos sociais. A ordem de importância das manifestações e seus fatores de relevância indicam um princípio para estas abordagens.

7.3 LIMITAÇÕES DA PESQUISA

Apesar dos objetivos propostos nesta pesquisa terem sido plenamente atingidos, certamente algumas limitações podem ser assumidas. A primeira diz respeito ao uso da amostragem não probabilística porque não permite grandes generalizações para outros setores da sociedade.

Decorrente desta, outra limitação diz respeito à amostragem por conveniência. Obviamente, os resultados dizem respeito à usuários acadêmicos e, portanto, os sistemas utilizados recuperavam objetos informacionais com as características impostas por estes usuários. No entanto, o mesmo instrumento de dados poder ser utilizado em outro enfoque modificando-se alguns critérios de relevância. Pode-se também citar, o efeito de clustering nos resultados dos fatores de relevância devido o tamanho da amostra.

7.4 SUGESTÕES PARA PESQUISAS FUTURAS

Como sugestões de futuras pesquisas propõe-se:

- Estudo de avaliação de desempenho de um SRI baseando-se nestes fatores de relevância;
- Levantamento de novos critérios de relevância provenientes da definição das estratégias de busca;
- Adaptação e utilização do instrumento de pesquisa em outros setores da sociedade como, por exemplo, órgãos governamentais e empresas privadas.

REFERÊNCIAS BIBLIOGRÁFICAS

BAEZA-YATES, R.; RIBEIRO-NETO, B. **Modern Information Retrieval**. New York: ACM Press, 1999.

BARRY, C.L. User-Defined Relevance Criteria: an Exploratory Study. **Journal of the American Society for Information Science**, v.45 (3), p. 149-159, 1994.

BARRY, C.L; SCHAMBER,L. Users' Criteria for Relevance Evaluation: a cross-situational comparison. **Information Processing and Management**, v.34 (2/3), p. 219-236, 1998.

BELKIN, N.J. Intelligent Information Retrieval: whose intelligence? In: **Proceedings of The Fifth international Symposium for Information Science**. Konstanz: Universtaeg Konstanz, p. 25-31, 1996.

BELKIN, N.J. **Interaction with Texts: information retrieval as information-seeking behavior**. Disponível em:< <http://citeseer.nj.nec.com/belkin93interaction.html>>. Acesso em: 16/09/2003.

BORLUND, P. Experimental Components for the Evaluation of Interactive Information retrieval systems. **Journal of Documentation**, v. 56(1), p., 2000.

BORLUND, P. The Concept of Relevance in IR. **Journal of the American Society for Information Science**, v. 54(10), p. 913-925, 2003.

BORLUND,P.; INGWERSEN, P. The Development of a Method for the Evaluation of Interactive Information Retrieval Systems. **Journal of Documentation**, v. 53, p. 225-220, 1997.

BRAJNIK, G.; MIZZARO, S.; TASSO, C. Evaluating User Interfaces to Information Retrieval Systems: a case study on user support. In: **19th International Conference on Research and Developments in Information Retrieval (SIGIR)**, P. 128-136, 1996.

BRUCE, H.W. A Cognitive View of The Situational Dynamism of User-Centered Relevance Estimation. **Journal of the American Society for Information Science**, v.45 (3), p. 142-148, 1994.

COOL, C.; BELKIN, N.; FRIEDER, O.; KANTOR,P. Characteristics of Texts Affecting Relevance Judgments. In M.E. Williams (Ed.), **Proceedings of the 14th National Online Meeting** . Medford, nJ: Learned Information, p. 77-84, 1993.

COSIJN, E.; INGWERSEN, P. Dimension of Relevance. **Information Processing and Management**, v.36, p.533-550, 2000.

DILLON, W.; MADDEN, T.; FIRTLE, N. **Marketing Research in a Marketing Environment**. St. Louis: Times Mirror/Mosby College, 1987.

DRAPER, S. W.; DUNLOP, M.D. **New IR-New Evaluation: the impact of interaction and multimedia on information retrieval and its evaluation**, 2002. Disponível em: <http://www.cs.strath.ac.uk/~mdd/research/publications/nrhm/new_IR_new_eval.pdf>. Acesso em : 30/01/2003.

DUNLOP, M.D.; JOHNSON, C.W.; REID, J. Exploring the Layers of Information Retrieval Evaluation. **Interacting with Computers**, v.10, p. 225-236, 1998.

EISENBERG, M.B. Measuring Relevance Judgments. **Information Processing and Management**, v. 24(4), p. 373-389, 1998.

FROEHLICH, T.J. Relevance Reconsidered-Towards an Agenda for the 21st Century : introduction to special topic issue on relevance research. **Journal of the American Society for Information Science**, v.45 (3), p. 124-134, 1994.

GLUCK, M. Exploring the Relationship Between User Satisfaction and Relevance in Information Systems. **Information Processing and Management**, v. 32(1), p. 89-104, 1996.

GREISDORF, H. Relevance Thresholds: a multi-stage predictive model of how users evaluate information. **Information Processing and Management**, v.39(3), p.403-423, 2002.

HERSH, W. Relevance and Retrieval Evaluation: perspectives from medicine. **Journal of the American Society for Information Science**, v.45 (3), p. 201-206, 1994.

HOWARD, D.L. Pertinence as Reflected in Personal Constructs. **Journal of the American Society for Information Science**, v.45 (3), p. 172-185, 1994.

INGWERSEN, P. **Information Retrieval Interaction**, 1992 Disponível em: <<http://www.db.dk/pi/iri>> Acesso em: 28/08/2003.

INGWERSEN, P. Cognitive Perspectives of Information Retrieval Interaction: elements of a cognitive IR theory. **Journal of Documentation**, v. 52(1), p. 3-50, 1996.

JANES, J.W. Relevance Judgments and the Incremental Presentation of Document Representation. **Information Processing and Management**, v.27 (6), p. 629-646, 1991.

JANES, J.W. Other People's Judgments: a comparison of users' and others' judgments of document relevance, topicality, and utility. **Journal of the American Society for Information Science**, v.45 (3), p. 160-171, 1994.

KEKÄLÄINEN, J.; JÄRVELIN, K. Evaluating Information Retrieval Systems under the Challenges of Interaction and Multidimensional Dynamic Relevance. **Proceedings of the 4th COLIS Conference**, p. 253-270, 2002.

KOWALSKI, GERALD. **Information Retrieval Systems: theory and implementation**. Boston: Kluwer Academic, 1997.

- MAGLAUGHLIN, K.L; SONNENWALD, D.H. User Perspectives on Relevance Criteria: a comparison among relevant, partially relevant, and not-relevant judgments. **Journal of the American Society for Information Science**, v.53 (5), p. 327-342, 2002.
- MALHOTRA, N.K. **Pesquisa de Marketing: uma orientação aplicada**. 3ª ed. Porto Alegre: Bookman, 2001.
- MATTAR, F.N. **Pesquisa de marketing: metodologia e planejamento**. São Paulo: Atlas, 1996.
- MIZZARO, S. How Many Relevances in Information Retrieval? **Interacting with Computers**, v.10, p. 303-320, 1998.
- PAO, M.L. Term and Citation Retrieval: a field study. **Information Processing and Management**, v. 29, p. 95-112, 1993.
- PARK, T.K. The Nature of Relevance in Information Retrieval: an empirical study. **The Library Quarterly**, v. 63, p. 135-141, 1993.
- PARK, T.K. Toward a Theory of User-Based Relevance: a call for a new paradigm of inquiry. **Journal of the American Society for Information Science**, v.45 (3), p. 135-185, 1994.
- PRESSMAN, R.S. **Engenharia de Software**. São Paulo: McGraw-Hill Ltda, 1995.
- REGAZZI, J.J. Performance Measures for Information Retrieval Systems: an experimental approach. **Journal of the American Society for Information Science**, v. 39 (4), p. 235-251, 1998.
- SALTON, G. The State of Retrieval System Evaluation. **Information Processing and Management**, v.28 (4), p.441-449, 1992.
- SAMARA, B.S; BARROS, J.C. **Pesquisa de Marketing**. 2ª ed. São Paulo: Makron Books, 1997.
- SARACEVIC, T. The Concept of “Relevance” in Information Science: a historical review, In T. Saracevic (Ed.), **Introduction to Information Science**, p. 111-151, New York : R.R. Bowker, 1970.
- SARACEVIC, T. Relevance: a review of and a framework for the thinking on the notion in information science. **Journal of the American Society for Information Science**, v. 39 (4), p. 235-251, 1975.
- SARACEVIC, T; KANTOR, P.; CHAMIS, A.Y; TRIVISON, D. A Study of Information Seeking and Retrieving I. Background and Methodology . **Journal of the American Society for Information Science and Technology**, v. 39, p. 161-176, 1988.
- SARACEVIC, T. Evaluation of Evaluation in Information Retrieval. In: **17th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval**, 1995.
- SARACEVIC, T. Modeling Interaction in Information Retrieval (IR): a review and proposal. In: **Proceedings of The American Society for Information Science**, v. 33, p. 3-9, 1996a.

SARACEVIC, T. Relevance Reconsidered. In: Information science: Integration in Perspectives. **Proceedings of The Second Conference on Conceptions of Library and Information Science (COLIS 2)**, p. 201-218, October, 1996b.

SARACEVIC, T. The Stratified Model of Information Retrieval Interaction: extension and application. **Proceedings of the American Society for Information Science**, v. 34, p. 313-327, 1997.

SARACEVIC, T.; SPINK, A.; WU, M.M. Users and Intermediaries in Interactive Information Retrieval: what are they talk about? **Proceedings of UM97: The 6th International Conference on User Modelling**, v. 383, p. 43-54, 1997.

SCHAMBER, L. Relevance and Information Behavior. **Annual Review of Information Science and Technology**, v.(29), p. 3-48, 1994.

SCHAMBER, L. Users' Criteria for Evaluation in a Multimedia Environment. **Proceedings of the American Society for Information Science**, Washington , DC, p. 126-133. Medford, NJ: Learned Information, Inc, 1994

SCHAMBER, L.; EISENBERG, M.; NILAN, M. A Re-examination of Relevance: toward a dynamic, situational definition. **Information Processing and Management**, v.26(6), p.755-776, 1990.

SMITHSON, J. Information Retrieval Evaluation in Practice: a case study approach. **Information Processing and Management**, v. 30(2), p. 205-221, 1994.

SPINK, A.; GREISDORF, H.; BATEMAN, J. From Highly to not Relevant: examining different regions of relevance. **Information Processing and Management**, v.34(5), p.599-621, 1998.

SPINK, A.; SARACEVIC, T. Human-computer Interaction in Information Retrieval: nature and manifestation of feedback. **Interacting with Computers**, v(10), p. 249-267, 1998.

SPINK, A. Information-Seeking and Mediated Searching. Part 1: theoretical framework and research design. **Journal of the American Society for Information Science**, v.53 (9), p. 695-703, 2002.

SPINK, A.; GREISDORF, H. Regions and Levels: measuring and mapping users' relevance judgments. 3^a ed. Porto Alegre: Bookman, 2001.

SU, L.T. Evaluation Measures for Interactive Information Retrieval. **Information Processing and Management**, v.28(4), p.503-516, 1992.

SU, L.T. The Relevance of Recall and Precision in User Evaluation. **Journal of the American Society for Information Science**, v.45 (3), p. 207-217, 1994.

SCHAMBER, L.; EISENBERG, M.; NILAN, M. A Re-examination of Relevance: toward a dynamic, situational definition. **Information Processing and Management**, v.26(6), p.755-776, 1990.

TANG,R.; SHAW,W.M.; VEVEA, J.L. Towards the Identification of the Optimal Number of Relevance Categories. **Journal of the American Society for Information Science and Technology**, v. 50(3), p. 254-264, 1999.

YAO, Y.Y. Information Retrieval Support Systems. Disponível em
<<http://suez.cs.gsu.edu/~cscyqz/research/wcci2002-cwi/Yao.pdf>> Acesso em: 10/03/2003.

ANEXOS

ANEXO A: INSTRUMENTO DE COLETA DE DADOS

Parte 2 – Avaliação de Agrupamentos e Critérios de Relevância

Gostaríamos que o Sr. (a) julgasse cada critério a seguir de acordo com o grau de relevância do mesmo. Para isto, utilize a escala em que **1= baixa relevância** (o documento recuperado não apresentou a qualidade descrita no critério) e **5 = alta relevância** (o documento apresentou completamente a qualidade descrita pelo critério).

A) Considerando a qualidade do “Tópico/ Assunto” apresentado no documento pesquisado atribua um grau de relevância para os critérios abaixo quanto	Baixa Relevância					Alta Relevância
	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
1. À concordância entre os termos da consulta efetuada e o assunto apresentado pelo documento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
2. À informatividade do <i>abstract</i> para sua decisão de aprofundar-se no texto completo.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
3. À concordância entre as informações apresentadas no documento e sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
4. À profundidade/ foco do conteúdo no documento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
5. À atualidade do assunto apresentado.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
6. Considerando os critérios acima, qual o grau de relevância do “Tópico/ Assunto” que o documento pesquisado apresentou?	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	

B) Considerando a qualidade da “Estrutura” do documento pesquisado atribua um grau de relevância para os critérios abaixo quanto	Baixa Relevância					Alta Relevância
	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
7. À clareza apresentada no documento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
8. À credibilidade (informações corretas e válidas) provida pelo documento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
9. À efetividade do método ou procedimento utilizado pelo autor.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
10. À facilidade de obtenção do documento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	
11. Considerando os critérios acima, qual o grau de relevância da “Estrutura” que o documento pesquisado apresentou?	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>	

C) Considerando a qualidade “Cognitiva” transmitida pelo documento pesquisado atribua um grau de relevância para os critérios abaixo quanto	Baixa Relevância		Alta Relevância		
12. Ao aumento de conhecimento acerca de sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
13. À obtenção de novas informações sobre o assunto pesquisado.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
14. À ampliação de novas perspectivas de conhecimento.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
15. Ao conhecimento adquirido sobre o assunto pesquisado.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
16. Ao reforço ou suporte ao seu ponto de vista sobre o assunto pesquisado.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
17. Considerando os critérios acima, qual o grau de relevância do aspecto “Cognitivo” que o documento pesquisado apresentou?	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>

D) Considerando a “Utilidade” do documento pesquisado atribua um grau de relevância para os critérios abaixo quanto	Baixa Relevância		Alta Relevância		
18. À utilidade para sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
19. À apresentação de novidades úteis para sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
20. À utilidade na resolução do problema informacional que você tem em mãos.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
21. À utilidade para tomar uma decisão sobre a necessidade de informação necessária.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
22. À utilidade para reduzir alguma incerteza quanto à sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
23. Considerando os critérios acima, qual o grau de relevância da “Utilidade” que o documento pesquisado apresentou?	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>

E) Considerando o aspecto “Motivacional” do documento pesquisado atribua um grau de relevância para os critérios abaixo quanto	Baixa Relevância Alta Relevância				
24. Ao estímulo que o conteúdo do documento lhe ofereceu a aprofundar-se mais no assunto.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
25. À inspiração a pesquisar outros objetivos além de sua necessidade de informação.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
26. À motivação a procurar pesquisas feitas pelo mesmo autor.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
27. À motivação a procurar os artigos citados pelo autor.	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>
28. Considerando os critérios acima, qual o grau de relevância do aspecto “Motivacional” que o documento pesquisado apresentou?	1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>

29 – De acordo com todo o julgamento feito anteriormente especifique abaixo qual o espaço de relevância em que se encontra o documento recuperado como um todo:

Baixa Relevância						Alta Relevância
1. <input type="checkbox"/>	2. <input type="checkbox"/>	3. <input type="checkbox"/>	4. <input type="checkbox"/>	5. <input type="checkbox"/>		

30 – Caso não especificado acima cite outros critérios que você utilizou para julgar o documento:

31 – Ferramenta de busca utilizada :

Proquest Science Direct Google Ebsco SSRN ACM Outro

Se outro qual? _____

Parte 3 – Informações de controle

Nome:	Telefone:
Nome do Artigo:	Data:
E-mail:	

**ANEXO B: VALORES DE *BETA* E SIGNIFICÂNCIA DOS DADOS
DAS TABELAS DE 13 A 17**

Tabela 19 - Dados estatísticos (*ex ante*) da tabela 13

		<i>Ex ante</i> (Grau de relevância)									
		1		2		3		4		5	
Critérios tabela 16		Beta	Sig	Beta	Sig	Beta	Sig	Beta	Sig	Beta	Sig
Critério 1		0,032	0,933	0,133	0,250	0,260	0,009	0,326	0,001	0,339	0,030
Critério 2		0,266	0,366	0,249	0,025	0,117	0,235	0,208	0,005	0,133	0,371
Critério 3		0,737	0,102	0,141	0,264	0,456	0,000	0,292	0,002	0,216	0,170
Critério 4		0,698	0,271	0,381	0,001	0,159	0,107	0,126	0,120	0,104	0,492
Critério 5		-0,433	0,479	0,327	0,008	0,281	0,004	0,237	0,004	0,273	0,072
Valores de R		0,929		0,872		0,747		0,743		0,550	

Tabela 20 - Dados estatísticos (*ex post*) da tabela 13

		<i>Ex post</i> (Grau de relevância)									
		1		2		3		4		5	
Critérios tabela 16		Beta	Sig	Beta	Sig	Beta	Sig	Beta	Sig	Beta	Sig
Critério 1		0,763	0,041	0,048	0,749	0,184	0,095	0,256	0,004	0,192	0,141
Critério 2		0,228	0,256	0,416	0,009	0,301	0,010	0,114	0,159	0,286	0,032
Critério 3		0,096	0,724	0,348	0,014	0,405	0,000	0,201	0,023	0,347	0,017
Critério 4		-0,173	0,481	0,192	0,191	0,380	0,000	0,522	0,000	0,169	0,260
Critério 5		-0,014	0,943	0,103	0,558	-0,027	0,772	0,151	0,054	0,212	0,137
Valores de R		0,898		0,615		0,820		0,789		0,682	

Tabela 21 - Dados de estatísticos (*ex ante*) da tabela 14

Critérios tabela 17	<i>Ex ante</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 7	0,355	0,149	0,382	0,002	0,244	0,003	0,457	0,000	0,362	0,000
Critério 8	0,633	0,118	0,236	0,033	0,301	0,002	0,332	0,000	0,396	0,000
Critério 9	0,186	0,583	0,343	0,006	0,479	0,000	0,343	0,000	0,318	0,000
Critério 10	-0,148	0,487	0,147	0,112	0,064	0,373	0,121	0,019	0,242	0,003
Valores de R	0,935		0,885		0,864		0,888		0,910	

Tabela 22 - Dados estatísticos (*ex post*) da tabela 14

Critérios tabela 17	<i>Ex post</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 7	0,273	0,233	0,340	0,000	0,153	0,032	0,449	0,000	0,370	0,000
Critério 8	0,540	0,147	0,358	0,002	0,503	0,000	0,306	0,000	0,356	0,004
Critério 9	0,274	0,348	0,363	0,002	0,385	0,000	0,351	0,000	0,427	0,000
Critério 10	-0,103	0,595	0,036	0,658	0,089	0,169	0,094	0,173	0,170	0,039
Valores de R	0,862		0,863		0,904		0,851		0,900	

Tabela 23 - Dados estatísticos (*ex ante*) da tabela 15

Critérios tabela 18	<i>Ex ante</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 12	0,477	0,084	0,037	0,721	0,263	0,018	0,246	0,001	0,179	,159
Critério 13	0,563	0,036	0,557	0,001	0,114	0,264	0,252	0,001	0,340	,016
Critério 14	0,288	0,224	0,170	0,253	0,400	0,000	0,181	0,015	0,185	,170
Critério 15	-0,142	0,478	0,169	0,151	0,077	0,455	0,185	0,016	0,371	,016
Critério 16	0,474	0,059	0,172	0,069	0,347	0,000	0,320	0,000	0,000	,998
Valores de R	0,970		0,875		0,793		0,893		0,825	

Tabela 24 - Dados de estatísticos (*ex post*) da tabela 15

Critérios tabela 18	<i>Ex post</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 12	-0,088	0,108	0,148	0,089	0,275	0,004	0,203	0,058	0,544	0,000
Critério 13	0,074	0,125	0,517	0,000	0,289	0,002	0,128	0,172	0,319	0,016
Critério 14	0,024	0,575	0,205	0,029	0,385	0,000	0,310	0,002	0,274	0,004
Critério 15	0,513	0,000	-0,05	0,574	0,152	0,096	0,259	0,016	-0,139	0,356
Critério 16	0,648	0,000	0,368	0,000	0,113	0,120	0,245	0,005	0,159	0,070
Valores de R	0,993		0,892		0,880		0,784		0,921	

Tabela 25 - Dados de estatísticos (*ex ante*) da tabela 16

Critérios tabela 19	<i>Ex ante</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 18	-	-	0,269	0,016	0,196	0,058	0,210	0,029	0,265	0,093
Critério 19	-	-	0,279	0,008	0,143	0,172	0,089	0,317	0,295	0,083
Critério 20	-	-	0,388	0,004	0,219	0,056	0,467	0,000	0,192	0,188
Critério 21	-	-	0,139	0,268	0,359	0,004	-0,021	0,832	0,131	0,316
Critério 22	-	-	0,139	0,242	0,116	0,290	0,173	0,067	0,115	0,366
Valores de R	-		0,871		0,783		0,681		0,779	

Tabela 26 - Dados de estatísticos (*ex post*) da tabela 16

Critérios tabela 19	<i>Ex post</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 18	-	-	0,272	0,049	0,343	0,008	0,334	0,000	-0,180	0,225
Critério 19	0,000	1,000	0,121	0,370	0,123	0,315	0,260	0,001	0,021	0,886
Critério 20	0,000	1,000	0,109	0,428	0,000	0,999	0,312	0,001	0,526	0,003
Critério 21	0,681	0,141	0,308	0,060	0,300	0,051	-0,066	0,509	0,190	0,359
Critério 22	0,000	1,000	0,188	0,240	0,246	0,065	0,271	0,004	0,040	0,841
Valores de R	0,681		0,776		0,758		0,842		0,634	

Tabela 27 - Dados estatísticos (*ex ante*) da tabela 17

Critérios tabela 20	<i>Ex ante</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 24	0,676	-	0,508	0,002	0,158	0,149	0,344	0,000	0,266	0,020
Critério 25	-0,294	-	0,136	0,386	0,475	0,000	0,252	0,001	0,208	0,084
Critério 26	-	-	0,242	0,063	0,092	0,405	0,323	0,000	0,063	0,574
Critério 27	0,882	-	0,233	0,051	0,218	0,054	0,329	0,000	0,704	,000
Valores de R	1,000		0,825		0,683		0,786		0,813	

Tabela 28 - Dados de estatísticos (*ex post*) da tabela 17

Critérios tabela 20	<i>Ex ante</i> (Grau de relevância)									
	1		2		3		4		5	
	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>	<i>Beta</i>	<i>Sig</i>
Critério 24	-0,082	0,556	0,350	0,004	0,349	0,002	0,214	0,004	0,402	0,001
Critério 25	0,902	0,000	0,204	0,081	0,254	0,022	0,190	0,014	0,144	0,211
Critério 26	-	-	0,318	0,003	0,202	0,056	0,223	0,013	0,355	0,005
Critério 27	0,087	0,538	0,319	0,002	0,302	0,003	0,468	0,000	0,259	0,041
Valores de R	0,900		0,780		0,781		0,870		0,779	