

UNIVERSIDADE FEDERAL DE SANTA CATARINA - UFSC
DEPARTAMENTO DE ENGENHARIA DE PRODUÇÃO E SISTEMAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO

**UM MODELO BASEADO EM AGENTES PARA O
PROCESSO DE COLETA DE DADOS VIA WEB**

ÁLVARO BORGES DE OLIVEIRA

Tese apresentada ao Curso de Pós-graduação em Engenharia de Produção da
Universidade Federal de Santa Catarina, para obtenção do título de Doutor em
Engenharia.

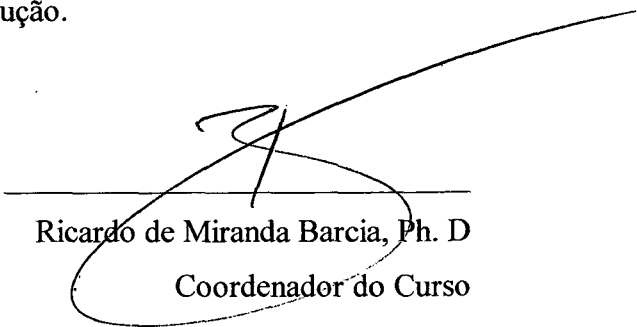
Prof. Rogério Cid Bastos, Dr.
Orientador

Florianópolis, fevereiro de 2000.

**UM MODELO BASEADO EM AGENTES PARA O PROCESSO DE COLETA
DE DADOS VIA WEB**

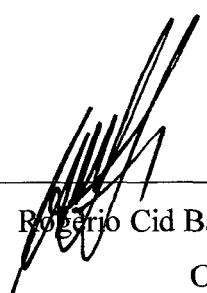
ÁLVARO BORGES DE OLIVEIRA

Esta Tese foi julgada adequada para a obtenção do título de Doutor em Engenharia, Especialidade Engenharia de Produção, e aprovada em sua forma final pelo Programa de Pós-graduação em Engenharia de Produção.



Ricardo de Miranda Barcia, Ph. D
Coordenador do Curso


BANCA EXAMINADORA



Rogério Cid Bastos, Dr.
Orientador

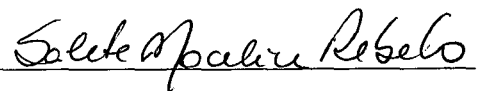


José Erno Taglieber, Dr.



Luiz Alberto Costa, Dr.

Raimundo Ghisoni Teivi, Dr.



Salete Mocelin Rebelo, Dra.

"A ciência apresenta raízes amargas e frutos doces".

Aristóteles

Dedico este trabalho aos meus filhos.

AGRADECIMENTOS

Agradeço:

Ao Grande Arquiteto do Universo;

À CAPES pelo apoio financeiro;

A minha companheira, Ilze Maria Tavares, que tanto me incentivou;

Aos meus filhos, Thessaly, Tharnier e Thamna, e meus pais, Juarez e Emília, pela compreensão de tantas ausências, no período em que me dediquei a este trabalho;

Aos meus irmãos e cunhados (as), em especial a minha irmã e amiga Gládis pelo incentivo que sempre me dedicou;

À Professora Ligia Ghisi, pelo incentivo e estímulo;

À Professora Sueli Petry da Luz, pela amizade, pela compreensão e pelo apoio dado a mim, o qual refletiu muito na conclusão deste trabalho;

Ao amigo Ronald Gomes, por parte da programação do sistema, e que tanto aprendemos junto;

Ao recente amigo, Professor Sérgio S. Januário, pelo apoio metodológico incomensurável;

Aos professores e funcionários da Engenharia de Produção e Sistemas;

Por fim, exalto o amigo de outrora, hoje meu orientador, Rogério Cid Bastos, ao qual faltariam adjetivos para elogiá-lo pelo acompanhamento neste trabalho, sem o qual não seria possível a realização.

RESUMO

Este trabalho é relativo ao desenvolvimento de um sistema aplicado a processos de coleta de dados em caráter diacrônico. Três estruturas de reflexão aplicada foram consideradas: o processo de pesquisa através de questionários; desenvolvimento em ambiente distribuído; utilização de Agentes, Raciocínio Baseado em Caso e Multicritério Difuso, gerando **Um Modelo Baseado em Agentes Para o Processo de Coleta de Dados Via Web**.

Para se alcançar tais propósitos, parte-se de uma crítica aos modelos tradicionalmente aplicados em processos de desenvolvimento de pesquisas quantitativas e qualitativas realizadas com aplicação de questionários por pesquisadores de campo bem como por meio de redes interligadas internet.

A construção da variável tempo é determinante neste trabalho. Ao invés de considerá-lo em perspectiva sincrônica, apresentamos aplicabilidade da pesquisa em sentido diacrônico com total habilitação do pesquisador sobre o intervalo de tempo e o prolongamento do período de coleta de dados. Desta forma, durante a aplicação da pesquisa o pesquisador dispõe de instrumentos de medidas como a confiabilidade, fidedignidade e a validade.

Uma das vantagens apresentadas é a capacidade de autogerenciamento e dimensionamento dos processos de coleta de dados, com racionalização do tempo e diminuição de custos por parte de empresas que buscam a projeção de serviços e produtos no mercado.

ABSTRACT

This paper concerns the development of a system applied to the processes of data collection in a diachronic matter. Three structures of applied discussion were considered: the research process through questionnaires, development of distributed environment, usage of Agents, Case Based Reasoning and Fuzzy multi-criteria, generating **A Model Based on Agents For The Process of Data Collection Via Web.**

To reach such purposes, it starts from a criticism to the traditionally applied models in processes of development of quantitative and qualitative researches performed with the application of questionnaires, by field researchers as well as through the interconnected networks.

The construction of the time variable is determined in this paper. Instead of considering it in a synchronic perspective, it is presented the applicability of the research in a diachronic sense with total habilitation of the researcher about the time gap and the stretching of the data collection period. This way, during the application of the research, the researcher has measurement instruments and reliability, loyalty and validity.

One of the presented advantages is the self-management skill and the dimensioning of the data collection processes, with time rationalization and the decreasing of expenses by the enterprises that look for the projection of its services and products in the market.

SUMÁRIO

<u>LISTA DE FIGURAS</u>	xi
<u>LISTA DE TABELAS</u>	xii
<u>LISTA DE TABELAS</u>	xii
<u>APRESENTAÇÃO</u>	13
<u>CAPÍTULO I</u>	15
<u>FUNDAMENTOS DE PESQUISA</u>	15
<u>1.1 Introdução</u>	15
<u>1.2 Métodos de Aplicação de Instrumentos de Pesquisa</u>	18
<u>1.3 Teoria da Confiabilidade e Fidedignidade</u>	21
<u>1.4 Tipos de confiabilidade</u>	23
<u>1.5 Teoria da Validade</u>	24
<u>1.6 Tipos de Validade</u>	25
<u>1.7 Confiabilidade e Validade</u>	30
<u>CAPÍTULO II</u>	34
<u>INTELIGÊNCIA ARTIFICIAL EM PESQUISA</u>	34
<u>2.1 Introdução</u>	34
<u>2.2 Pesquisas via E-mail versus Correio</u>	35
<u>2.3 Projeto e Métodos de Pesquisa via E-mail</u>	40
<u>2.4 Pesquisas e Técnicas de IA</u>	41
<u>CAPÍTULO III</u>	44
<u>MODELO TEÓRICO DE SISTEMA DE PESQUISA</u>	44
<u>3.1 Introdução</u>	44
<u>3.2 Implementação do Sistema de Pesquisas</u>	47
<u>3.2.1 Público Alvo</u>	49
<u>3.2.2 Critérios</u>	50
<u>3.2.3 Perguntas</u>	50
<u>3.2.4 Âncoras</u>	51
<u>3.3 Métodos de Seleção de Perguntas</u>	51
<u>3.3.1 Método de Multicritério Difuso</u>	52

3.3.1.1 Resolução do Problema.....	53
3.3.2 Método Aleatório.....	56
3.4 Arquitetura de Agentes Para o Sistema Proposto	57
3.4.1 Agentes da Classe Pesquisador	59
3.4.1.1 Agente de Interface.....	59
3.4.1.2 Agente de Publicação	60
3.4.1.3 Agente Busca IP.....	60
3.4.1.4 Agente Busca Pesquisa.....	60
3.4.1.5 Agente Busca Pergunta	61
3.4.1.6 Agente RBC Pergunta.....	61
3.4.2 Agentes da Classe Pesquisas	62
3.4.2.1 Agente RBC Pesquisa	62
3.4.2.2 Agente Pesquisado	63
CAPÍTULO IV	64
IMPLEMENTAÇÃO DO SISTEMA PROPOSTO.....	64
4.1 Tecnologia Empregada	64
4.2 Sistema de Pesquisa e Publicação	64
4.3 Análise dos Dados	67
4.4 Agentes de Busca	69
4.5 Sistema na Web.....	70
CONCLUSÕES E RECOMENDAÇÕES.....	73
REFERÊNCIAS BIBLIOGRÁFICAS.....	76
BIBLIOGRAFIA	82
APÊNDICE I.....	86
TEORIA DE AGENTES.....	86
1.1 Introdução	86
1.2 Histórico.....	86
1.3 Conceitos de Agentes	88
1.4 Atributos dos Agentes	89
1.5 Taxionomia dos Agentes	91
1.6 Arquiteturas de Agentes.....	93
1.6.1 Linguagem de Comunicação de Agentes.....	94
APÊNDICE II.....	95
RACIOCÍNIO BASEADO EM CASOS.....	95

<u>1.1 Introdução</u>	95
<u>1.2 Histórico</u>	95
<u>1.3 Conceitos de RBC</u>	96
<u>1.4 Fundamentação de RBC</u>	97
<u>1.4.1 Representação de Casos</u>	98
<u>1.4.2 Recuperação de Casos</u>	101
<u>1.4.3. Reutilização de Casos</u>	103
<u>1.4.4. Revisão do Caso</u>	104
<u>1.4.5. Retenção do Caso (Aprendizagem)</u>	104
<u>APÊNDICE III</u>	106
<u>Tomada de Decisão Difusa</u>	106
<u>1.1 Discussão Geral</u>	106
<u>1.2 Tomada de Decisão Multicritério</u>	108
<u>1.2.1 Aplicação de Kauffmann</u>	108
<u>1.2.1.1 Exemplo</u>	111
<u>1.2.2 Aplicação de Klir</u>	114

LISTA DE FIGURAS

<u>Figura 1. Dois campos envolvidos na pesquisa</u>	26
<u>Figura 2. Validade acumulativa</u>	29
<u>Figura 3. Metáfora do alvo</u>	31
<u>Figura 4. Relação entre confiabilidade e validade</u>	32
<u>Figura 5. Sistema de pesquisa proposto</u>	45
<u>Figura 6. Arquitetura do sistema proposto</u>	46
<u>Figura 7. Elaboração, Aplicação e Análise</u>	48
<u>Figura 8. Escolha de perguntas</u>	52
<u>Figura 9. Conjunto suporte para pergunta</u>	54
<u>Figura 10. Classes de agentes</u>	58
<u>Figura 11. Tela principal do sistema</u>	65
<u>Figura 12. Tela perguntas da pesquisa</u>	66
<u>Figura 13. Tela de publicação da pesquisa</u>	66
<u>Figura 14. Tela agente</u>	67
<u>Figura 15. Tela salvar dados da pesquisa</u>	68
<u>Figura 15. Tela resultado em percentagem</u>	68
<u>Figura 16. Tela agente de busca pesquisa</u>	69
<u>Figura 17. Tela de agente de busca pergunta</u>	70
<u>Figura 18. Tela Inicial na Internet</u>	71
<u>Figura 19. Escolha da pesquisa a ser respondida</u>	71
<u>Figura 20. Respondendo a pesquisa</u>	72
<u>Figura 21. Sistema Antigo versus Sistema Proposto</u>	73

LISTA DE TABELAS

<u>Tabela 1. Diferença entre os tipos de pesquisa</u>	20
<u>Tabela 2. Estudos anteriores comparando correio e e-mail</u>	36
<u>Tabela 3. Exemplo de uma pesquisa</u>	52
<u>Tabela 4. Conjunto pertinência para cada pergunta</u>	55
<u>Tabela 5. Rank das Perguntas</u>	56
<u>Tabela 6. Exemplo de iterações</u>	57
<u>Tabela 7. Rank das perguntas</u>	57

APRESENTAÇÃO

As inovações tecnológicas são vistas como a forma mais eficaz de economizar níveis de trabalho, simplificando rotinas complexas. Aumentam a produtividade e a quantidade de informações existentes.

Contudo, criam-se paradoxos e conflitos que exigem novas soluções. Uma observação interessante diz respeito ao tempo de resposta de sistemas informatizados. Quanto maior a velocidade, menor deverá ser o tempo de resposta, pois do contrário agrega-se elevados níveis de estresse e ansiedade.

O mesmo se processa com as informações que circulam na rede Internet. Assim, o que se revela interessante é como e quanto a tecnologia irá auxiliar na ampliação do negócio, aumentando os lucros e/ou reduzindo custos.

Como objetivo geral deste trabalho, propõe-se construir um sistema de pesquisa para coletar, armazenar e analisar informações em ambientes distribuídos, utilizando alta tecnologia e agregação de técnicas de Inteligência Artificial.

Têm-se como objetivos específicos:

- Desenvolver um sistema capaz de gerenciar pesquisas;
- Utilizar conjuntos difusos na elaboração de instrumentos de coleta de dados;
- Desenvolver protótipos de agentes inteligentes para coletar pesquisas já elaboradas, perguntas e publicar pesquisas;
- Agregar o ambiente Web no processo decisório;

As inovações criadas a partir do uso das tecnologias de informação, têm alterado e produzido formas de negócios e de agregar valor.

A economia da era digital apresenta duas virtudes de grande interesse: o comércio eletrônico e a indústria da tecnologia da informação.

Por comércio eletrônico entende-se o processo de negócios que realizam as transações na rede Internet ou em algum outro sistema proprietário baseado em redes de computador. A indústria da tecnologia de informação tem apresentado um crescimento tão rápido que muitos dos produtos existentes, não eram previsíveis anos atrás.

Sistemas que oferecem subsídios para estas duas vertentes são de interesse da Engenharia de Produção.

A importância do presente trabalho tem fulcro nos fatos de propor o uso de técnicas como Lógica Difusa, Raciocínio Baseado em Casos e Agentes Inteligentes para coleta e prospecção de informações em um sistema baseado em redes de computadores.

Este trabalho está estruturado em 4 capítulos. No primeiro capítulo é feito um estudo dos fundamentos da pesquisa. Esse estudo engloba os aspectos pertinentes ao entendimento do sistema que se elaborou, bem como a sua importância em relação aos métodos de medidas como confiabilidade e validade de uma pesquisa.

O segundo capítulo disserta um estudo empregado na aplicação de métodos de pesquisa via e-mail e correio, por ser estes os métodos mais utilizados atualmente, segundo Couper (1999). Diante do apresentado, contribui-se com uma teoria de pesquisas em Inteligência Artificial.

No terceiro capítulo é apresentada a proposta do sistema, onde é descrito o funcionamento e as tecnologias utilizadas para o desenvolvimento do mesmo.

O quarto capítulo descreve o sistema propriamente dito, mostrando o funcionamento do que foi implementado.

Por fim, há três apêndices que dissertam sobre Agentes, Raciocínio Baseado em Caso e Multicritério Difuso, teorias que foram utilizadas na elaboração do sistema.

CAPÍTULO I

FUNDAMENTOS DE PESQUISA

1.1 Introdução

A orientação metodológica de qualquer pesquisa de caráter científico tem, por dever, a necessidade de apresentar-se diante do campo teórico-metodológico do modo mais sofisticado. Basicamente, há três correntes clássicas de interpretação de fenômenos sociais na análise científica. Cada um desses métodos de abordagem – Materialismo Histórico Dialético, Funcionalismo e Sociologia Interpretativa ou Compreensiva – implica em reflexões epistemológicas de acentuado grau de abstração.

Nestes termos, a ciência é considerada como uma das formas de produção de conhecimento, e não única e, nem mesmo, a verdadeira. O que se tem como problematização científica são os fundamentos sobre a construção dos “objetos de pesquisa”.

Em caráter interpretativo, a construção do objeto se concede pela racionalidade. Em outras palavras, o conhecimento não está no objeto em si, senão na construção racional de concepções sobre os objetos. Concebendo-se que a racionalidade é limitada em termos de conseguir abordar todos os elementos, variáveis e categorias existentes na natureza própria de cada objeto de pesquisa, corroboram-se as concepções relativistas sobre o processo de construção de conhecimento científico.

Desta forma, o método científico existe mesmo para garantir objetividade do conhecimento, isto é, deve garantir formas unívocas (ou situações de pesquisa próximas disso), para que qualquer cientista possa seguir o mesmo caminho (ou método utilizados) em condições muito semelhante de pesquisa (contextos cultural, social, político, econômico, ambiental, etc.) e chegar a resultados iguais ou diferentes (não necessariamente opostos, mas apenas diferentes).

Assim, procede-se ao desenvolvimento do conhecimento científico, aprimorando-se método e aspectos puramente teóricos de análise científica.

Todavia, método não se restringe puramente aos procedimentos de pesquisa, mas aos aspectos teóricos de abordagem. Nenhum conteúdo teórico, portanto, estará desprovido de referências empíricas, nem tampouco os processos indutivos de análise estarão desprovidos de embasamento teórico.

As concepções de realidade que se tem por referência, nos dias atuais (relação a tempo de existência e uso de tecnologias avançadas que rompem com concepções tradicionais de espaço e tempo – este fenômeno é extremamente recente como percepção de realidade) avançam para determinações de espaço e tempo superando as construções de realidade física.

Este novo contexto implica em determinações, segundo referências, para ações quotidianas determinadas por este contexto.¹

A internet contribui, efetivamente, para a produção de mudança de hábitos e referências de ação quotidiana e de planejamento de futuro. De outro modo, passa a contribuir para mudanças fundamentais da organização social em termos de filosofia de vida (como a construímos no quotidiano) e em termos de filosofia de mundo (o que fazemos hoje que contempla a idéia de como desejamos que seja o futuro). Assim, como formas de diferenciação de organização da sociedade, a internet constitui-se como instituição social, ou seja, como referência de valor para a constituição da ação quotidiana e de planejamento de devir.

Neste panorama de investigação, cabe à ciência a análise e proposições de recursos capazes de compreensão das formas de organização da sociedade a partir de modelos institucionais de referência a partir da internet.

Vários processos de trabalho estão disponíveis na construção social deste “novo século” baseada em Tecnologia da Informação (reconstrução de espaço e tempo,

¹ Este pode ser considerado um dos elementos que marcam a emergência do século XXI, superando inclusive a padronização cronológica de tempo. De outro modo, não são as percepções de cronos que passa a determinar o período de um século, senão os fenômenos sociais que determinam modos diferentes de relacionamentos sociais (interpessoais e entre pessoas mediados por objetos, tal qual as tecnologias avançadas).

por meio de emissão e recepção de informações em tempo real e espaço físico desnecessário como referência).

Um deles, ao qual consideramos ser extremamente importante pelo fato de permitir a expansão em escalas amplas de uso (não afirmamos aqui quaisquer configurações de democracia) são as ferramentas que potencializam instrumentalmente a tecnologia de informação como referência quotidiana. Ilustrativamente, podemos afirmar, sem quaisquer prejuízos analíticos, os procedimentos de consultas de obras e dados em instituições sólidas em nosso contexto de existência, como bibliotecas, fenômenos políticos de instituições de representação de poder, comentários individuais sobre assuntos diversos, acesso à mídia escrita, à mídia eletrônica e à mídia radiofônica, etc.

A ampliação de uso da Tecnologia da Informação não significa o acesso a computadores. Estes são, tão somente, bases instrumentais de atividades; compõem-se de elementos físicos que não diferem do modelo de ação referencial de quaisquer tecnologia existentes no século XX. Ou seja, um automóvel é, tecnologicamente, tão avançado quanto um computador; deste ponto de análise não expressam qualquer diferença e, por conseqüência, estão matizados no mesmo campo, o que nos permite afirmar que a tecnologia da informação são os procedimentos e processos de acesso às informações. Aqui, de fato, entramos em outro campo analítico, qual seja, o de construir software aplicativos que, fundamentalmente, constituem-se no caminho de referência da tecnologia da informação. Este objeto de análise é, para a parte da ciência que se dedica a tais preocupações, um dos elementos chave para crítica e, se for o caso, intervenção social.

O que passamos a desenvolver, neste campo, implica em posicionamento analítico que se refere não somente ao conjunto de instrumentos em si, senão ao conjunto de efeitos e produções sociais que são determinados por estes processos e, ao mesmo tempo, a sua existência enquanto referência para as atividades quotidianas. Qualquer *persona* que se vincula ao espaço público da sociedade está, mesmo que indiretamente vinculada à instituição social que marca o “novo século”: a Tecnologia da Informação.

Estas reflexões estão, portanto, fundamentadas teoricamente e, por concepção de ciência aqui apreciada, deve servir empiricamente como vetor componente da instituição social. Esta última parte refere-se, propriamente, a processos de pesquisa na WEB.

A Pesquisa na WEB apresenta-se, até então, imerso num paradigma sincrônico de procedimento (como sintonizar uma estação rádio). De outro modo, cada proposição de uso na tecnologia de informação concebe-se a si próprio como mais um elemento disponível (mais um software).

1.2 Métodos de Aplicação de Instrumentos de Pesquisa

Os instrumentos de pesquisa comumente aplicados são questionários e roteiros de entrevistas, doravante definidos como método de questionário e entrevista.

A aplicação de questionários e a realização de entrevistas, nos últimos anos, sofreram profundas mudanças graças aos avanços ocorridos na Tecnologia de Informação. Estes foram automatizadas através de pesquisas por telefone usando discagem randômica, de quiosques computadorizados em lugares públicos, de e-mail, etc. Também apareceram com mais freqüências em nossas vidas como num quarto de hotel sobre o criado-mudo, num restaurante com garçom apresentando uma pesquisa sobre satisfação do cliente, numa loja ao comprarmos um eletrodoméstico, e recentemente num Web site.

Os questionários foram por muito tempo aplicada via correio, direcionada a grupos previamente estabelecidos como numa empresa, ou o pesquisador levava o instrumento de porta em porta ou em hasta pública. Quando aplicada via correio trás consigo a característica de fácil aplicabilidade e administração da pesquisa. Por outro lado tem suas desvantagem também, pois, a taxa de respostas é baixa. Estudos mais aprofundados sobre o assunto pode ser visto em Couper *et al* (1999), Dillman (1991), Ramos (1998) e Tse *et al* (1995).

As entrevistas apresentam-se, comumente, através de pesquisas pessoais, onde o entrevistador trabalha diretamente com o entrevistado ou por telefone, neste caso se pode coletar as informações mais rapidamente. Este método não é o escopo deste trabalho. Uma discussão sobre o assunto pode ser encontrada em Aquilino (1994), Aquilino (1990), Bradburn (1983), Leenw (1988) e Gfroerer (1991).

Um paralelo entre questionário e entrevista é mostrado na Tabela 1, apresentado por Trochim (2000). Observa-se também que, na Tabela 1, cada método de pesquisa tem suas vantagens e desvantagem, no entanto, deve-se observar alguns fatores que influenciam diretamente na escolha do método, como questões sobre a população alvo, a amostra da população, as perguntas, o objeto de estudo, o preconceito e a administração, que adquire o seguinte significado no presente estudo:

- *População alvo*: alguns fatores devem ser considerados ao se escolher o método de pesquisa empregado, como: a acessibilidade; se esta pode ser numerada; se há analfabetos ou não; qual o grau de linguagem a ser empregado; se está disponível e cooperará e, questão geográfica.
- *Amostra*: é o grupo de pessoas que se terá contato de alguma maneira. Existem várias questões a serem consideradas como a sua disponibilidade, onde encontrar, quem é o pesquisado, levar em consideração se toda a população pode ser amostrada e se a taxa de resposta é um problema para sua pesquisa.
- *Perguntas*: ressaltam a natureza do que se quer perguntar aos pesquisados, determinando o método. Os pontos a serem considerados são: o tipo de pergunta, quão complexo pode ser, o tamanho da pergunta, a escala a ser utilizada e a ordem que se apresentarão.
- *Objeto de estudo*: é o tema a ser levantado na pesquisa. Na escolha do método de pesquisa deve-se levar em consideração se o pesquisado deve entender do assunto e se sua aplicação requer consulta em livros ou registros por parte do pesquisado.

Assuntos	Questionnaire			Interview	
	Grupo	E-mail	Correio	Pessoal	Fone
São possíveis apresentações visuais?	Sim	Sim	Sim	Sim	Não
São categoriais de respostas a longo prazo, possíveis?	Sim	Sim	Sim	???	Não
É a privacidade uma característica?	Não	Sim	Não	Sim	???
O método é flexível?	Não	Não	Não	Sim	Sim
São viáveis questões abertas?	Não	Não	Não	Sim	Sim
São leitura e escrita necessárias?	???	Sim	Sim	Não	Não
Você pode julgar a qualidade da resposta?	Sim	Não	???	Sim	???
São possíveis altos níveis de resposta?	Sim	Não	Sim	Sim	Não
Você pode explicar o resultado do estudo na pessoa?	Sim	Não	Sim	Sim	???
O custo é baixo?	Yes	Yes	Não	Não	Não
Are Staff & Facilities Needs Low?	Sim	Sim	Não	Não	Não
Does It Give Access to Dispersed Samples?	Não	Sim	Não	Não	Não
Does Respondent Have Time to Formulate Answers?	Não	Sim	Sim	Não	Não
Is There Personal Contact?	Sim	Não	Sim	Sim	Não
Is A Long Survey Feasible?	Não	Não	Não	Sim	Não
Is There Quick Turnaround?	Não	Sim	Não	Não	Sim

Tabela 1. Diferença entre os tipos de pesquisa, adaptado de (Trochim 2000).

- *Preconceito*: trás à tona relação íntima do pesquisado. Questões como os anseios sociais e perguntas respondidas falsamente podem ser evitados e devem ser levados em considerações.
- *Administração*: aqui se considera a viabilidade da pesquisa, levando-se em consideração os custos, as facilidade de aplicação, tempo e pessoal envolvido.

Além das questões levantadas, Trochim (2000) considera outros fatores na construção do instrumento de pesquisa como considerar o conteúdo da pergunta, o fraseado da pergunta, o formato da resposta e a seqüência das perguntas. Pode-se ver mais detalhes em Trochim (2000).

A construção de um instrumento de pesquisa é complexo, pois objetiva construir uma ferramenta de medida. Nota-se que pesquisa em Ciências Sociais diferem das pesquisas feitas em campos como Física e Química por muitas razões. Uma razão é que em Ciências Sociais, a exemplo, a psicologia, está tentando medir a intangibilidade das coisas, como atitudes, comportamentos, emoções, personalidades, etc. Na física usa-se uma régua para medir a distância e em química um cilindro graduado para medir o volume, sendo que em ciências sociais não podemos colocar emoções ou fenômenos sociais dentro de um cilindro graduado ou usar uma régua para medir a atitude de alguém ou de um grupo social.

Como resultado os cientistas sociais têm desenvolvido seus próprios significados de medida para atitudes, comportamentos, emoções, personalidades, etc. A dificuldade encontrada, na aplicação de qualquer um dos métodos diz respeito à consistência e periodicidade da aplicação do instrumento e que nunca se sabe, com certeza, se se está medindo o que se quer. Dessa forma, o pesquisador em Ciências Sociais dispõe de bons instrumentos de medidas: *Reliability* (confiabilidade) e *Validity* (validade).

Quando se elabora uma pesquisa, é importante que as medições sejam tanto confiáveis como válidas. Se não, seus instrumentos são basicamente inúteis e diminuem-se as chances de medir com precisão aquilo que se pretende medir.

Assim, é importante determinar como se irá medir um fenômeno em particular. Este processo de medição é importante porque permite saber se estamos no caminho certo e se está realmente medindo aquilo que se deseja medir. Ambas a confiabilidade e a validade são essenciais em uma boa medição, pois são as primeiras defesas contra a formação de conclusões imprecisas (isto é, aceitação ou rejeição incorreta de sua hipótese de pesquisa).

1.3 Teoria da Confiabilidade e Fidedignidade

Estamos familiarizados com termos como consistência, previsibilidade, segurança, estabilidade e frequência, pois é termos que vêm as mentes quando falamos sobre confiabilidade.

Mais amplamente definida, a confiabilidade de uma medição refere-se à consistência ou frequência da medição de algum fenômeno. Se uma medição é um instrumento confiável, isto significa que o instrumento pode medir a mesma coisa mais de uma vez ou utilizar mais de um método e produzir o mesmo resultado.

Quando falamos de confiabilidade, não estamos falando de indivíduos, estamos na verdade falando de pontuações.

A pontuação observada é um dos principais componentes da confiabilidade, isto é, a pontuação que se observa em um cenário de pesquisa, sendo composta de uma *pontuação verdadeira* e uma *pontuação errada*.

A pontuação verdadeira é um conceito teórico, pois não existe uma maneira de se saber qual é o verdadeiro. A pontuação verdadeira reflete o valor verdadeiro de uma variável.

A pontuação errada é a razão pela qual a pontuação observada é diferente da verdadeira. A pontuação errada é ainda dividida em erros de método (ou sistêmico) e erros de traço (ou aleatório).

O erro de método refere-se a qualquer coisa que causa uma diferença entre a pontuação observada e a pontuação verdadeira devido à situação de teste. Por exemplo, qualquer tipo de interrupção (música alta, vozes, trânsito) que ocorre enquanto os estudantes estão realizando um teste pode causar a distração do mesmo e pode afetar suas notas no teste.

Por outro lado, o erro de traço é causado por quaisquer fatores relacionados às características da pessoa que está realizando o teste que podem aleatoriamente afetar a mediação. Um exemplo de um erro de traço no trabalho é quando indivíduos estão cansados, com fome ou desmotivados.

Estas características podem afetar seu desempenho em um teste, fazendo com que as pontuações pareçam piores do que seriam se os indivíduos estivessem atentos, bem alimentados ou motivados.

Desta forma a confiabilidade pode ser vista como a razão entre a pontuação verdadeira e a pontuação verdadeira mais à pontuação errada, ou:

$$\frac{\textit{Pontuação Verdadeira}}{\textit{Pontuação Verdadeira} + \textit{Pontuação Errada}}$$

Definida a confiabilidade e quais são seus componentes, provavelmente imagina-se como obter confiabilidade. Colocado de forma simples, o grau de confiabilidade pode ser aumentado pelo decréscimo da pontuação errada. Assim, se se quer um instrumento confiável, precisa-se diminuir o erro.

Como declarado anteriormente, nunca se pode saber a pontuação real de uma mediação. Portanto, é importante observar que a confiabilidade não pode ser calculada; ela só pode ser estimada. A melhor maneira para estimar a confiabilidade é medir o grau de correlação entre as diferentes formas de mediação. Quanto maior a correlação, maior a confiabilidade.

Dentro deste contexto, precisa-se observar três principais aspectos da confiabilidade: equivalência, estabilidade e homogeneidade.

A equivalência refere-se ao grau de concordância entre duas ou mais medições feitas aproximadamente ao mesmo tempo. Para que ocorra a estabilidade, deve ser feita uma distinção entre a frequência da medição e a do fenômeno que está sendo medido. Isto é obtido empregando dois indicadores. Por fim, a homogeneidade trata da avaliação de o quão bem os diferentes itens em uma medida parecem refletir a característica que se está tentando medir. Aqui a ênfase está nos relacionamentos internos, ou na consistência interna.

1.4 Tipos de confiabilidade

Apresenta-se a confiabilidade sob três tipos. O primeiro tipo de confiabilidade é a confiabilidade das formas paralelas. Esta é uma medida de equivalência e envolvem a administração de duas formas diferentes ao mesmo grupo de pessoas e a obtenção da correlação entre as duas formas. Quanto maior a correlação entre as duas formas, mais equivalentes às formas.

O segundo tipo de confiabilidade, a confiabilidade teste-reteste, é uma medida de estabilidade que examina a confiabilidade com o passar do tempo. A maneira mais fácil de medir a estabilidade é administrar o mesmo teste em dois momentos diferentes (usando o mesmo grupo de pessoas) e obter uma correlação entre os dois testes. O problema com a confiabilidade do re-teste é o tempo de espera entre os dois testes. Quanto mais você espera, menor sua estimativa de confiabilidade.

Finalmente, o terceiro tipo é a confiabilidade interindicadores, uma medida de homogeneidade. Com a confiabilidade interindicadores duas pessoas qualificam um comportamento, objeto ou fenômeno, e determinam o grau de concordância entre eles. Para determinar a confiabilidade interindicadores pega-se o número de concordâncias e divide pelo número total de observações.

1.5 Teoria da Validade

Ao começarmos a descrever validade temos que perguntar: “validade do quê?”. Quando pensamos em validade em pesquisa, pensa-se nos componentes da pesquisa. Poderíamos dizer que uma medida é um componente válido, ou que uma amostra foi tirada, ou que o *design* teve uma grande valia. Mas todos estes preceitos estão tecnicamente incorretos. Medidas, amostras e *designs* não têm validade, somente proposições podem ser ditas válidas.

Tecnicamente, deveríamos dizer que uma medida leva a conclusões válidas ou que uma amostra possibilita inferências válidas, e assim por diante. É uma proposição, inferência ou conclusão que podem “ter” validade.

Fazemos muitas inferências ou conclusões diferentes enquanto elaboramos uma pesquisa. Muitos destes estão relacionados ao processo de pesquisar e não são as hipóteses principais do estudo. Muito embora, como os tijolos que vão à construção de uma parede, este processo intermediário e proposições metodológicas promovem a fundação para conclusões substantivas as quais desejamos referir-nos. Por tal razão, virtualmente todas as pesquisas sociais envolvem medidas e observação. E, em todo momento que medimos ou observamos estamos preocupados se estamos medindo o que

queremos medir ou como nossas observações são influenciadas pelas circunstâncias dentro das quais elas são feitas.

Alcançamos conclusões sobre a qualidade de nossas medidas – conclusões que terão um importante papel ao referirmo-nos a pontos substantivos mais amplos, de nosso estudo. Quando falamos sobre a validade de pesquisa, geralmente nos referimos as muitas conclusões que chegamos sobre a qualidade de diferentes partes de nossa pesquisa metodológica. Para Trochim (2000) é a melhor aproximação disponível à verdade de uma dada proposição, inferência, ou conclusão.

1.6 Tipos de Validade

Validade subdivide-se em quatro tipos: conclusão, interna, construção e externa. Cada tipo refere-se a uma questão metodológica específica. Para entender os tipos de validade, devemos saber algo sobre como investigamos uma pesquisa. Por que todos os quatro tipos de validade são realmente operacionais quando do estudo de questões causais e para tal usaremos um estudo causal para estabelecer o contexto.

A Figura 1 mostra que há na realidade dois campos que estão envolvidos na pesquisa, segundo Trochim (2000).

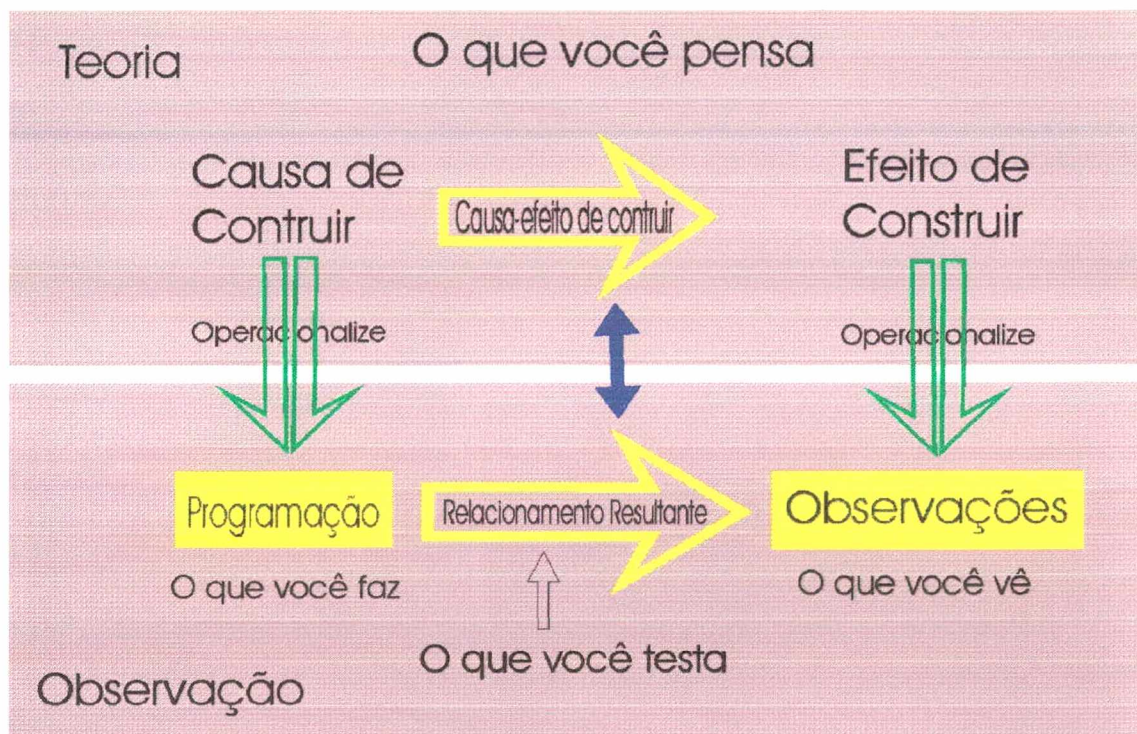


Figura 1. Dois campos envolvidos na pesquisa adaptado de Trochim (2000).

O primeiro, em cima, é o campo da teoria. É o que acontece dentro de nossas mentes como pesquisadores. É onde mantemos nossas teorias sobre como o mundo opera.

O segundo, abaixo, é o campo das observações. É o mundo real, no qual traduzimos nossas idéias, nossos programas, tratamentos, medidas e observações.

Quando conduzimos uma pesquisa, estamos continuamente pairando sobre um e outro desses dois campos, entre o que pensamos do mundo e o que está acontecendo nele. Quando estamos investigando uma relação de causa e efeito, temos uma teoria (implícita ou de outra forma), de qual é a causa (construção da causa).

Portanto, se estamos testando um novo programa educacional, temos uma idéia de como ele pareceria idealmente. O mesmo no lado do efeito, temos uma idéia do que estamos idealmente tentando afetar e medir (construção do efeito). Mas cada um destes, causa e efeito, tem que ser traduzido em coisas reais, em um programa ou tratamento e uma medida ou método observacional.

Usa-se o termo operacionalização para descrever o ato de traduzir uma construção em sua manifestação. Na verdade, é quando pegamos nossa idéia e a

descrevemos como uma série de operações ou procedimentos ou ao invés de ser somente uma idéia em nossas mentes, torna-se uma entidade pública que qualquer um pode olhar e examinar por si mesmo. Por exemplo, diz-se que gostaria de medir auto-estima (uma construção), mas quando você mostra uma medida de auto-estima de dez itens escritos em papel, que se desenvolve para este propósito, os outros podem olhar para ele e entender mais claramente o que você pretendeu dizer com o termo auto-estima.

Destarte, podemos voltar a explicar os quatro tipos de validade. Estes se montam uns sobre os outros, com dois deles (conclusão e interno) referindo ao campo da observação abaixo da Figura 1, um deles (construção) enfatizando as ligações entre o de baixo e o de cima, e o último (externo) sendo primariamente relacionado à variação acima.

Suponha que desejamos verificar se uma sala de aula virtual sob a Web melhora o entendimento do material do curso. Assuma que pegamos estas duas construções, a construção da causa (o “site” na Web) e o efeito (entendimento), e os operacionalizamos, tornando-os reais, isto é, construindo o Web site e uma medida de conhecimento do material do curso. Diante do problema exposto podemos considerar cada um dos tipos de validade.

Quanto à questão Validade da Conclusão, pergunta-se se neste estudo há uma relação entre as duas variáveis? No contexto do exemplo que estamos considerando, a questão pode ser feita: há uma relação entre o Web Site e o entendimento do material do curso? Há muitas conclusões ou inferências que podemos pegar para responder tal questão. Poderíamos, por exemplo, concluir que há uma relação. Poderíamos concluir que há uma relação positiva. Poderíamos inferir que não há relação.

Quanto à Validade interna, assumindo-se que há uma relação neste estudo, é a relação, uma relação causal? Somente porque encontramos que o uso do Web Site e conhecimento são correlatos, não podemos assumir necessariamente que o Web Site, usualmente causa o conhecimento. Ambos poderiam, por exemplo, ser causados pelos mesmos fatos. Assim, pode ser que estudantes com melhores condições financeiras, os quais tem maiores recursos, teria maior probabilidade de usar e ter acesso ao Web Site e seriam mais eficazes nos testes. Quando queremos argumentar que nosso programa ou

tratamento causou o resultado em nosso estudo, podemos considerar a validade interna de nossa argumentação causal.

Na Construção da validade, assumindo-se que há uma relação causal neste estudo, podemos argumentar que o programa refletiu bem nossa idéia da construção da medida? Em termos simples, implementamos o programa que pretendíamos implementar e medimos o resultado que queríamos medir? Em outros termos ainda, operacionalizamos bem as idéias de causa e efeito? Quando nossa pesquisa está pronta, gostaríamos de concluir que fizemos um bom trabalho de operacionalização de nossas construções, podemos acessar a validade da construção desta conclusão.

Para a validade externa, assumindo-se que há uma relação causal neste estudo entre as construções da causa e efeito, podemos generalizar este efeito para outras pessoas, lugares ou tempos? Tendemos argumentar que as descobertas de nossas pesquisas têm implicações para outros grupos e indivíduos em outros contextos e em outros tempos. Quando o fazemos, podemos examinar a validade externa destes argumentos.

Nota-se na Figura 2, conforme Trochim (2000), como a questão a que cada referência de tipo de validade pressupõe uma resposta afirmativa à sua anterior.

Pode-se concluir que os tipos de validade montam-se um sobre o outro. A Figura 2 ainda mostra a idéia de acúmulo como um lance de escadas, combinada com a questão chave de cada tipo de validade.

Para qualquer inferência ou conclusão, sempre há possíveis ameaças à validade, razões pelas quais a conclusão ou inferência pode estar errada. Idealmente, tenta-se reduzir a plausibilidade das mais prováveis ameaças à validade, deixando o mais plausível a conclusão alcançada no estudo.

Para exemplificar, imagine um estudo examinando se há uma relação entre a quantidade de treinamento em uma tecnologia específica e as variações subsequentes ao uso desta tecnologia. Porque o interesse está em uma relação, é considerado um ponto de validade conclusiva.

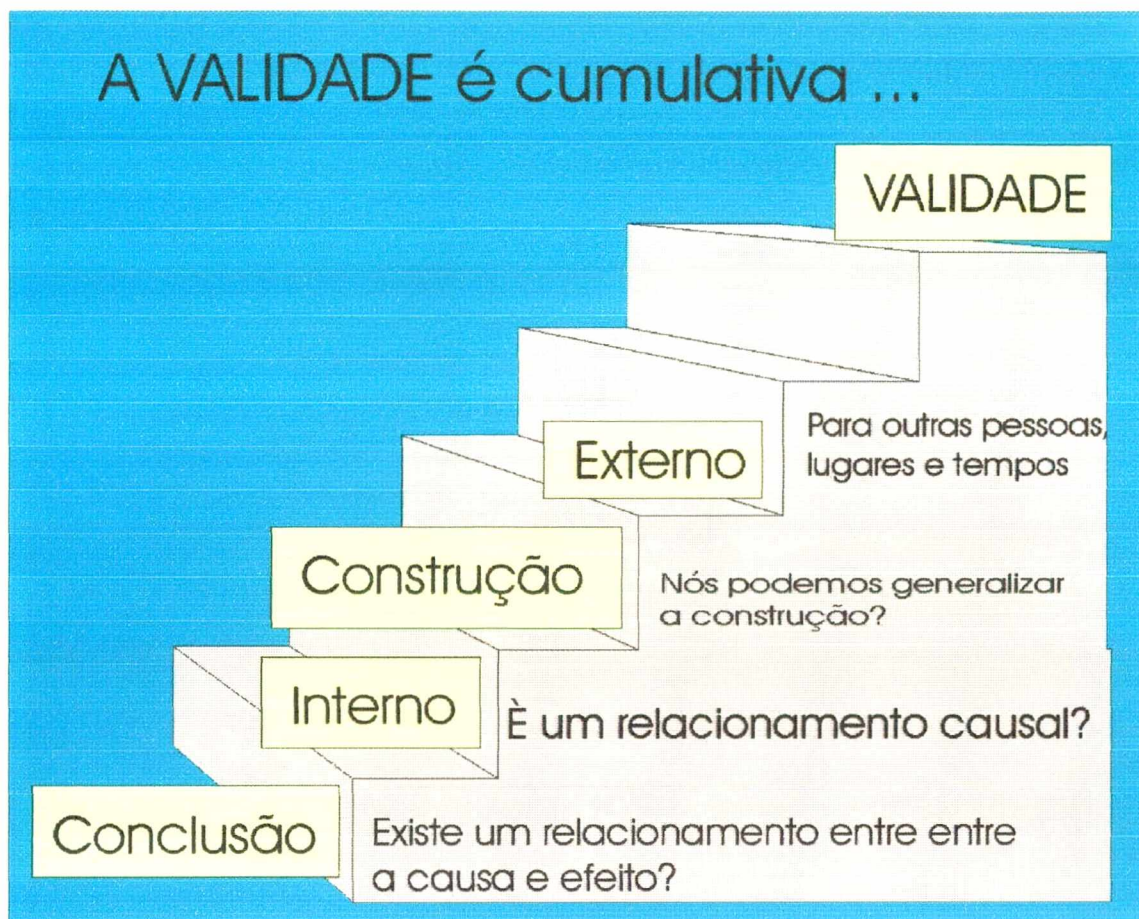


Figura 2. Validade acumulativa adaptado de Trochim (2000).

Por outro lado, assumo que o estudo foi completo e nenhuma correlação significativa entre a quantidade de treinamento e adoção de variações é encontrada. Nesta base é concluído que não há relação entre os dois. Como poderia esta conclusão estar errada, isto é, quais são as ameaças à validade? É possível que não haja poder estatístico suficiente para detectar uma relação mesmo que ela exista. Pode ser que o tamanho da amostra seja muito pequeno ou a medida da quantidade do treinamento não é confiável. Ou quem sabe, afirmações do teste correlativo são violadas em razão das variáveis usadas. Pode ser que havia irrelevâncias esporádicas no contexto do estudo ou heterogeneidade esporádica nos pesquisadores que aumentaram a variabilidade nos dados e tornaram mais difícil a visão da relação de interesse. A inferência de que não há relação será mais forte, tem maior validade conclusiva, se puder mostrar que estas explicações alternativas não são boas.

As distribuições podem ser examinadas para ver se estão de acordo com as afirmações do teste estatístico, ou análise conduzida para determinar se há suficiente poder estatístico.

A teoria da validade, e as muitas listas de ameaças específicas, nos proporcionam um útil esquema para acessar a qualidade das conclusões da pesquisa. A teoria é geral em finalidade e aplicabilidade, bem articulada em suas suposições filosóficas, tornando-se difícil explicar adequadamente em algumas páginas, assim, para um aprofundamento no assunto ver Colosi (2000), Durgan (2000), Mugo (2000) e Rymarchyk (2000).

O relacionamento entre a confiabilidade e a validade é simples de compreender: uma medição pode ser confiável, mas não ser válida. Entretanto, uma medição deve ser confiável antes de poder ser válida. Assim, a confiabilidade é uma condição necessária, mas não suficiente para a validade. Em outras palavras, uma medição pode avaliar de modo consistente um fenômeno (ou resultado), mas a menos que esta medição teste o que se deseja testar, ela não é válida.

1.7 Confiabilidade e Validade

Freqüentemente, pensamos em confiabilidade e validade como idéias separadas, mas, na realidade, está relacionada uma a outra. Há duas maneiras que se pode pensar sobre a relação entre confiabilidade e validade.

Para Trochim (2000) uma das metáforas a respeito de confiabilidade é comparar com um alvo, visto na Figura 3. Pensemos no centro do alvo como o conceito que se está tentando “medir”. Para cada pessoa que se está “medindo”, dá-se um tiro no alvo. Mede-se o conceito perfeitamente para uma pessoa se estamos atingindo o centro do alvo. Se não, estamos errando o centro. O mais “fora” que se estiver para aquela pessoa, o mais distante se está do centro.

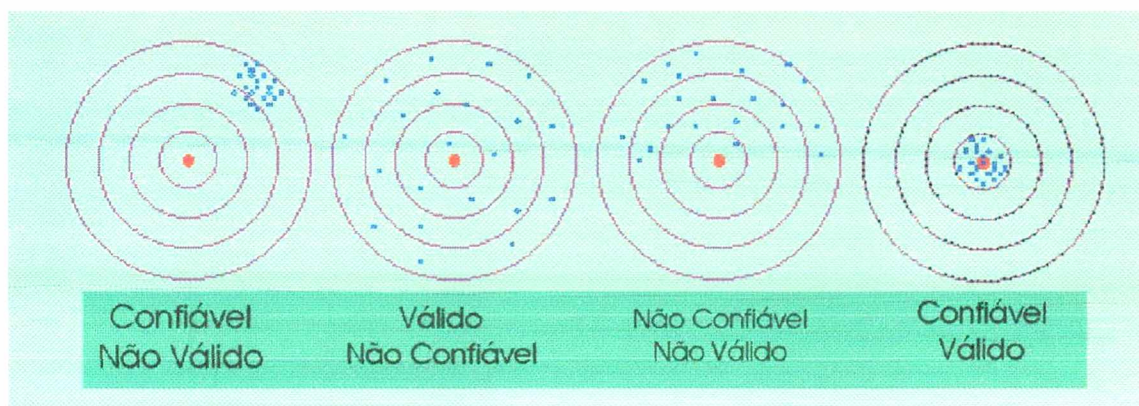


Figura 3. Metáfora do alvo adaptado de (Trochim 2000).

Na primeira, se está atingindo o alvo, consistentemente, mas se está errando o centro do alvo. Quer dizer, se está consistentemente e sistematicamente medindo o valor errado para todos os pesquisadores. Esta medida é confiável, mas não válida (quer dizer, é consistente, mas errada).

A segunda mostra tiros que estão furtivamente espalhados ao longo do alvo. Às vezes, atinge o centro do alvo, mas em média, consegue-se a resposta correta para o grupo (mas não muito bem para os indivíduos). Neste caso, consegue-se uma estimativa grupal válida, mas é inconsistente. Neste caso, pode-se ver claramente que confiabilidade está diretamente relacionada à variabilidade de sua medida.

O terceiro alvo mostra um caso onde seus tiros estão espalhados ao longo do alvo e se está consistentemente errando o centro. Sua medida, neste caso, não é nem confiável nem válida.

Finalmente, vemos o alvo ou cenário “Hobin Hood”, consistentemente atinge o centro do alvo. Sua medida é confiável e também válida.

Um outro modo em que podemos pensar sobre a relação entre confiabilidade e validade é mostrado na Figura 4, onde estabelecemos uma tabela 2x2. As colunas da tabela indicam que se está tentando medir o mesmo ou diferentes conceitos. As linhas mostram se está usando o mesmo ou diferentes métodos de medida. Suponhamos dois conceitos que gostaríamos de medir: a habilidade verbal e a matemática de estudantes.

Também suponhamos que vamos medir cada uma destas de duas maneiras. Primeiro pode-se usar um exame um escrito. Segundo podemos pedir à professora do

aluno, para dar-nos uma cotação ou média, da habilidade do aluno baseada em sua própria observação em sala de aula.

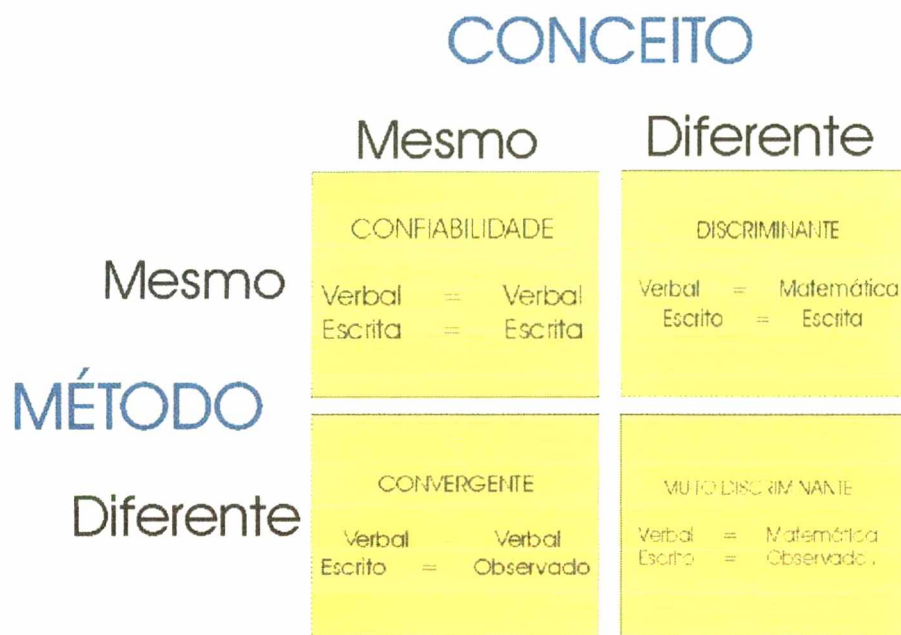


Figura 4. Relação entre confiabilidade e validade adaptado de (Trochim 2000).

A primeira célula do lado superior esquerdo mostra a comparação dos pontos obtidos com o teste de escrito e o teste escrito para habilidade verbal, mas como podemos medir ele com ele mesmo? Poderíamos fazer isto através da estimativa de confiabilidade do teste escrito através de uma correlação, re-teste, formas paralelas, ou uma medida da consistência interna (item 2.4). O que estamos estimando nesta célula é a confiabilidade da medida.

A célula no lado esquerda inferior mostra a comparação da medida do teste escrito, para habilidade verbal, com a observada pelo professor na sala de aula, também para a habilidade verbal. Porque estamos tentando medir o mesmo conceito, estamos olhando para uma validade convergente (item 2.6).

A célula superior direita mostra a comparação do teste escrito, para habilidade verbal, com o teste escrito, para habilidade matemática. Aqui, estamos comparando dois conceitos diferentes (habilidade verbal versus habilidade matemática) e, então, iríamos esperar que a relação fosse mais baixa que a comparação do mesmo conceito com ele mesmo (ex: habilidade verbal versus habilidade verbal ou habilidade matemática versus

habilidade matemática). Então, estamos tentando discriminar entre dois conceitos e consideraríamos este discriminante, validade.

Finalmente, temos a célula no lado inferior direito. Aqui, estamos comparando o teste escrito, para habilidade verbal, com a observada pelo professor na sala de aula, para a habilidade matemática. Como a célula superior à direita, estamos também, tentando comparar dois conceitos diferentes (habilidade verbal versus habilidade matemática), então esta é uma estimativa discriminante da validade. Mas, aqui, estamos também tentando comparar dois diferentes métodos de medida (teste escrito versus a observação do professor na sala de aula). Então, chamaremos isto muito discriminante, para indicar que esperaríamos a relação nesta célula ser mesmo menor que naquela acima.

Quando olhamos para a confiabilidade e validade desta maneira, vemos que, mais do que sendo distintos, na verdade formam um contínuo. De um lado está a situação onde os conceitos e métodos de medida são os mesmos (confiabilidade) e do outro está a situação onde conceitos e métodos de medida são diferentes (validade muito discriminante).

Diante do exposto, cabe-nos repensar os métodos de elaboração de pesquisa, em razão de deferência, da medida de determinada hipótese. O próximo capítulo faz um paralelo entre os métodos de aplicação de pesquisa tradicionais e os que utilizam a Internet (e-mail) e propõe novas técnicas, como a Inteligência Artificial, a fim de melhorar a elaboração de instrumentos de pesquisa.

CAPÍTULO II

INTELIGÊNCIA ARTIFICIAL EM PESQUISA

2.1 Introdução

O processo de conversão de pesquisas administradas por entrevistador ou usando abordagens baseadas em papel para métodos assistidos por computador (converting interviewer-administered surveys from paper-based approaches to computer assisted methods CATI and CAPI) está em estágio bastante avançado, sendo os Estados Unidos e Europa segundo Couper e Nicholls (1998) os precursores.

Até recentemente as pesquisas pelo correio não haviam sido afetadas por esta transição. Com a proliferação das comunicações eletrônicas nos últimos anos, pesquisas via correio eletrônico (e-mail) estão cada vez mais sendo promovidas como uma alternativa rápida e eficiente em relação ao custo para as pesquisas tradicionais via correio segundo Oppermann (1995) e Swoboda (1997). Schuldt e Totten (1994) afirmam que as pesquisas via e-mail podem tornar-se o método padrão de coleta de dados no século XXI.

Proponentes das pesquisas via e-mail em geral reconhecem que o acesso atualmente limitado ao e-mail por grande parte do público torna este método inadequado para pesquisas de população em geral.

Assim, pesquisas via e-mail tem sido bastante testadas em populações fechadas (tais como empregados em uma única organização), onde a penetração da tecnologia do e-mail é alta. Em tais casos, onde a cobertura não é preocupação, o e-mail é uma alternativa potencialmente atraente em comparação às pesquisas tradicional via correio.

Por outro lado se pode dizer que as pesquisas por e-mail são uma tecnologia de transição que irá levar ao desenvolvimento de pesquisas baseadas na Web, e que, por

consequente, há uma tendência de substituição dos métodos baseados em papel pelos eletrônicos.

Assim, é importante explorar a utilidade e eficiência deste método alternativo de coleta de dados, o qual é objeto deste trabalho.

2.2 Pesquisas via E-mail versus Correio

Há uma literatura pequena, porém crescente, comparando a coleta de dados via e-mail e a coleta via correio, ora visto que pouco se fala de pesquisa sobre a Web.

Um ponto central destas comparações tem sido as taxas relativas de respostas adquiridas usando estas abordagens alternativas. Uma Segunda área de interesse está relacionada aos ganhos de eficiência gerados pelo e-mail, ambos em termos de custos reduzidos e oportunidade ou velocidade de retorno. Vários estudos também examinaram as diferenças na qualidade dos dados obtidos pelos dois métodos, quer em termos de taxas de itens com dados faltosos ou em termos de distribuição de respostas substantivas. Questões de cobertura recebem atenção limitada, assumindo-se que o acesso ao e-mail seria um pré-requisito para a inclusão no estudo.

Os poucos experimentos de modo que foram realizados, segundo Couper et al (1999), variam em dimensões tão diferentes que é difícil fazer comparações diretas entre os mesmo. Também é difícil extrair tendências definitivas desta literatura restrita. Enquanto não há dúvidas de que as pesquisas por e-mail são mais rápidas do que as pesquisas via correio, e mais baratas em termos de custos de coleta de dados, as descobertas das taxas de respostas e na qualidade dos dados estão umas tanto misturadas.

Resumimos as descobertas sobre a taxa de resposta para cada um destes estudos brevemente na Tabela 2.

Estudo	Taxas de resposta		População	Tamanho da amostra e projeto
	Correio	E-mail		
Kiesler & Sproull (1986)	75%	67%	Usuários ativos de computadores na Carnegie-Mellon U	75 cada; chamada telefônica anterior a pesquisa para solicitar a participação; instrumento de 18 itens.
Parker (1992)	38%	68%	140 ex-empregados da AT&T	100 e-mail para todos aqueles com endereço de e-mail; correspondência via correio para aqueles sem endereço de e-mail; 32 itens de pesquisa.
Schuldt & Totten (1994)	57%	19%	Faculdades de MIS e Marketing em diretórios de associações	200 cartas via correio; 343 e-mail com endereços conhecidos (218 dos quais possuíam endereço usável); instrumento de 2 páginas.
Kittleson (1995)	77%	28%	Assinantes do diretório de e-mail de educadores da área da saúde	Todos os 153 receberam separadamente uma correspondência tradicional e um e-mail de 1 item sobre tópicos diferentes, no mesmo dia.
Tse et al. (1995)	27%	6%	Equipe de administração e ensino da Universidade Chinesa de Hong Kong com endereços de e-mail listados	200 cada; instrumento de 4 itens, sem acompanhamento.
Tse (1996)	52%	7%	Mesma acima	200 cartas, 300 e-mail, 1 lembrete; instrumento de 14 itens
Mehta & Sivadas (1995)	46%	40%	Usuários ativos de grupos BBS	60 e-mail, 192 cartas com endereços conhecidos, sem solicitação, instrumento de 5 páginas (300 linhas)
Bachman, Elfrink & Vazzana (1996)	66%	53%	Coordenadores e chefes de Faculdades de Administração	244 e-mail para todos aqueles com e-mail conhecido; 244 cartas para a amostra daqueles sem e-mail; instrumento de 2 páginas.
Comley (1997)	18%	9%	3700 nomes e endereços comprados de uma revista da Internet	1221 e-mail para todos aqueles com e-mail; correspondência para equilibrar; instrumento de 12 páginas
Schaefer & Dillman (1998)	58%	58%	Corpo Docente da Universidade Estadual de Washington	226 cada; instrumento de 46 itens

Tabela 2. Estudos anteriores comparando correio e e-mail.

Analisando a Tabela 2, observa-se que, primeiro, percebe-se que apenas três dos estudos, Kiesler e Sproull (1986), Tse et al. (1995), Tse (1996), Schaefer e Dillman (1998), escolheram aleatoriamente os indivíduos. Nos outros casos, aqueles com endereço de e-mail conhecido eram escolhidos para a pesquisa com e-mail e os outros para o correio tradicional.

Segundo apenas dois tipos de estudos (Parker, 1992; Schaefer e Dillman, 1998) encontraram taxas de resposta mais altas para o e-mail do que para o correio. Também se observa uma grande variação nas taxas de resposta em todos estes estudos, desde uma tão baixa como 18% até uma tão alta como 75% para o correio, e uma de 6% até uma de 67% para e-mail. Estas diferenças drásticas entre os estudos provavelmente refletem várias diferenças de projeto, desde a população escolhida, conteúdo e extensão do instrumento, até os métodos de transmissão e os números de contato. Evidentemente, são necessários mais estudos antes de esboçar conclusões definitivas sobre as diferenças nas taxas de resposta entre correio e e-mail.

Ao mesmo tempo em que vários autores têm especulado sobre as razões para o diferencial não-resposta em seus estudos, não houve tentativa sistemática para explorar os mecanismos subjacentes a estas diferenças. Sossdian e Sharp (1980) acreditam ser importante distinguir entre falha de acesso e resistência à resposta. As baixas taxas de resposta do e-mail em alguns estudos são causadas por problemas de transmissão ou por falta de vontade da pessoa pesquisada em completar uma pesquisa via e-mail?

Se a primeira for a razão, melhorias na tecnologia e cobertura do e-mail poderão reduzir a diferença entre o correio e o e-mail. Por outro lado, se há resistência diferencial do pesquisado, as preocupações sobre o erro de não-resposta podem ser ainda maiores. Se há uma falta de vontade maior de responder uma pesquisa por e-mail do que uma pesquisa equivalente via correio, isto se deve a dificuldades técnicas de leitura, preenchimento e retorno do instrumento.

Em termos de qualidade dos dados, os resultados também são mistos. Kiesler e Sproull (1986) descobriram que 22% dos pesquisados, por cartas, via correio não completaram ou anularam um ou mais itens, comparados a 10% dos pesquisados de e-mail. Schaeffer e Dillman (1998) descobriram que o e-mail era mais completo que a

carta via correio: 60,4% dos pesquisados de e-mail completaram pelo menos 95% dos itens, enquanto que apenas 56,6% dos pesquisados, via correio o fizeram.

Contudo, Bachman, Elfrink e Vazzana (1996) encontraram uma taxa de dado faltoso mais alta no e-mail (3,1%) do que no correio (0,7%). Tse et al. (1995) relatou tal taxa, levemente mais alta, porém não de modo significativo ($p > 0,05$), para e-mail (média de três itens faltosos em 36 itens) do que para a correspondência tradicional (média de 1,15 itens faltosos). Comley (1997) também encontrou taxas não significativas ($p > 0,1$) de omissão de item de 1,2% para e-mail e 0,4% para correio. Finalmente, Mehta e Sivadas (1995) não encontraram diferenças nas taxas de itens faltosos entre os modos.

Apenas dois estudos examinaram as respostas à perguntas abertas. Schaeffer e Dillman (1998) relataram a obtenção de respostas mais longas a questões abertas em e-mail (uma média de 3,9 linhas de texto) do que no correio (média de 1,7 linhas de texto). Bachman, Elfrink e Vazzana (1996) descobriram que 22% dos pesquisados de e-mail e 5% dos pesquisados, via correio forneceram uma resposta à única pergunta aberta.

Os resultados mistos sobre estas medidas de qualidade dos dados sugerem que se precisa analisar melhor as possíveis razões para estas diferenças. Se os instrumentos foram projetados para serem similares no formato, por que deveria-se esperar diferenças nas taxas de itens faltosos? Existem diferenças técnicas em termos de métodos de entrada que levam a diferentes taxas de preenchimento dos itens ou respostas a perguntas abertas? A natureza codificável do e-mail versus o formato em página da correspondência comum levam a diferenças por causar a inadvertida omissão de itens? Ou diferentes preocupações sobre segurança levam ao preenchimento de menos itens em um modo do que em outro? Sem maiores detalhes sobre cada um destes estudos não podemos responder estas questões, mas os resultados sugerem que é importante examinar as possíveis causas de tais diferenças sejam elas questões técnicas, diferenças de projeto ou formato, ou efeitos motivacionais associados ao modo.

Resumindo, então, ainda há muito a ser estudado sobre como (e por que) as pesquisa por e-mail podem diferir das pesquisas via correio, especialmente no que tange a não-resposta e a qualidade dos dados.

Além das comparações de modo explícitas até aqui discutidas, há vários relatórios recentes sobre pesquisas via e-mail que podem fornecer uma visão dos possíveis efeitos de um e-mail sobre erros de pesquisa. Opperman (1995) conduziu uma pesquisa via e-mail com membros da Associação Americana de Geógrafos com endereços de e-mail listados. Ele selecionou 665 pessoas, que produziram 500 endereços de e-mail usáveis. Ele obteve uma taxa de resposta de 49%, usando dois envios de mensagens via e-mail.

Swoboda et al. (1997) relatou os resultados de uma pesquisa de 9 itens enviada a endereços obtidos de vários grupos de notícias eletrônicas em Munique, Alemanha. Usando um projeto de um só envio, sem mensagens de aviso prévio ou acompanhamento posterior, eles enviaram um total de 8859 mensagens via e-mail, 5,3% das quais foram retornadas como não transmitidas. Eles relataram uma taxa de resposta de 20,4%. Além disso, para testar as mensagens não transmitidas, Swoboda et al. (1997) enviaram mensagens de e-mail a 118 endereços fictícios. Após 2 semanas, 91 haviam sido retornadas como não transmissíveis, deixando 27 sem serem considerados. Isto sugere que a consideração de mensagens não transmitidas pode subestimar as dificuldades técnicas de transmissão dos instrumentos.

Smith (1997) enviou uma pesquisa via e-mail de 66 itens para “webmasters” profissionais. Metade da amostra de 300 receberam uma pesquisa via e-mail direto, codificado em HTML, e a outra metade recebeu um pedido de participação, depois do que a pesquisa foi enviada. Apenas 11 dos 150 da amostra de e-mail direto retornaram o questionário, comparados a 42 no outro grupo que solicitaram a pesquisa (dos quais 20 realmente responderam a pesquisa). Smith também relatou ter experimentado vários problemas com falhas de transmissão e incompatibilidade de browser.

Conclui-se assim, que os resultados variados entre todos estes estudos sugerem que as questões de projeto e implementação podem ser importantes para pesquisas, de forma que novas tecnologias sejam agregadas.

Além disso, dado a provável proliferação de pesquisas via e-mail, é importante que estudos adicionais sobre comparação de modos sejam feitos sob várias condições. Vários dos estudos acima, também sugerem que muitas questões técnicas relacionadas às pesquisas via e-mail permaneceram sem solução, especialmente quando a pesquisa

não foi conduzida em um sistema fechado (uma única organização na qual todas as pessoas usam o mesmo sistema de e-mail).

Desta forma, este trabalho, vem dar sua contribuição, agregando técnicas de Inteligência Artificial, bem como propor um novo modo de prospecção.

2.3 Projeto e Métodos de Pesquisa via E-mail

Há três abordagens básicas ao desenvolvimento de pesquisa via e-mail. Uma abordagem produz um arquivo executável que é baixado pelo pesquisado e executado em sua máquina.

A vantagem desta abordagem é que uma interface gráfica de usuário (GUI) pode ser empregada para facilitar o preenchimento da pesquisa. O arquivo executável pode fazer total uso da assistência do computador para dar perfeição, variação, consistência e outros, e permite ramificações, espaços para preenchimento e outras características comuns às auto-entrevistas assistidas por computadores (CASI) segundo Couper e Nicholls, (1998).

Contudo, esta abordagem possui desvantagens como exigir que todos os usuários tenham sistemas compatíveis, e que se sintam confortáveis em baixar e executar arquivos.

A segunda abordagem alternativa transmite o questionário como um arquivo texto (ASCII) no corpo da mensagem do e-mail. O pesquisado então simplesmente desce pela mensagem recebida, respondendo aos itens, e retorna a mensagem ao remetente quando terminar.

Para Dillman (1978) e Dillman (1991) esta abordagem é um sistema bastante independente (a mensagem pode ser lida com qualquer combinação software/hardware) e ela é o equivalente mais próximo da pesquisa tradicional via correio, com todos os benefícios e desvantagens, tal o projeto que ela traz. Parece-nos que a maioria dos estudos sobre e-mail relatados acima (item 3. 2) adotou esta abordagem. Estas duas

abordagens representam filosofias diferentes sobre o que deveria ser uma pesquisa via e-mail: uma pesquisa CASI ou uma pesquisa eletrônica análoga à pesquisa via correio tradicional.

A terceira abordagem, descrita por Batagelj e Vehovar (1998) e Clayton e Weking (1998), é a que está ganhando popularidade. Esta abordagem usa o e-mail para convidar os pesquisados a entrar em um site seguro no qual completam a pesquisa. A pesquisa via Web pode ser projetada como uma alternativa ao instrumento CASI ou como uma pesquisa de e-mail navegável equivalente.

2.4 Pesquisas e Técnicas de IA

Visto nos itens anteriores que as pesquisas vem sendo aplicadas via e-mail, com tratamento muito semelhante às abordagens baseadas em papel, precisa-se repensar as etapas de elaboração de pesquisa, a fim de que esta tenha todos os atributos esperados como confiabilidade e validade.

É notório o desenvolvimento diário das tecnologias que envolvem a Internet. Conseqüentemente, todos os dias aparecem linguagens, utilitários, aplicativos, etc. para a tecnologia em questão. E, cada vez mais, técnicas de Inteligência Artificial vêm sendo aplicadas na construção das tecnologias que envolvem a Internet.

Desta forma, na construção, aplicação e recuperação de dados, se tratado como etapas a serem cumpridas e sempre conexas, a elaboração de programas aplicando técnicas de Inteligência Artificial são soluções que podem resultar na melhoria da qualidade das pesquisas.

Primeiramente, no que diz respeito à construção do instrumento de pesquisa devemos considerar as facilidades que os bancos de dados nos propõe. De ante mão podemos ter um banco de pergunta com milhões de perguntas direcionadas a públicos alvos, o que tornaria a elaboração da pesquisa mais fácil, pois já houve pesquisa a respeito do objeto a ser pesquisado e armazenado em forma de pergunta. Estamos falando aqui de um software que se propõe a elaborar os instrumentos e este deverá ser

aplicado via correio, e-mail ou Web, cabendo ao pesquisador a escolha do que lhe convier.

No entanto, já que temos a tecnologia a nosso favor, como a Internet, porque não aplicá-la, mas não mais de forma estática, isto é, através de arquivos .txt, softwares a serem baixados, *forms*, e sim utilizando recursos como multimídia, CGI, GUI, etc. Então, durante a construção do instrumento de pesquisa, o software em questão deve permitir utilizar os recursos no qual o ambiente a ser aplicado oferece, pois segundo Smith (1997) pesquisas iterativas, dinâmicas aumentam a motivação do pesquisado.

O software também deve ser flexível a ponto de nos permitir a escolha, por exemplo, do tipo de resposta: aberta, fechada ou aberta/fechada, uma data para que a pesquisa fique disponível na Web, ou ainda que determine o número de pesquisado a responder os instrumento que está sendo elaborado e os critérios relacionados ao público alvo. Se pensarmos, como dito anteriormente, em um banco de dados com milhões de perguntas, ao selecionarmos o público alvo este número se restringiria, pois as perguntas estariam relacionadas, também a outros públicos alvo. Outra diminuição da quantidade de perguntas seria a escolha do tipo de resposta, se aberta ou fechada. Mas, mesmo assim, ainda nos sobraria muitas perguntas o que seria um transtorno tanto para o pesquisado, para responder, como para o pesquisador que terá a função de selecioná-las.

A Lógica difusa (Apêndice 3) é uma técnica de IA a qual pode ser aplicada para ajudar o pesquisador a escolher, dentre as perguntas restantes, as que irão compor o instrumento de pesquisa. Especificamente, em Lógica Difusa temos o Multicritério Difuso que nos permite, através de critérios preestabelecidos ordenar de forma crescente ou decrescente as possíveis alternativas. Assim, se temos um rol de perguntas selecionadas e não queremos que todas participem do instrumento de pesquisa, podemos então estabelecer critérios com o público alvo, de forma a ser comparado, com os mesmos critérios, para cada pergunta arrolada.

Findo o processo de construção do instrumento de pesquisa, este deve ser publicado na Web para que possa ser respondido, estamos na segunda etapa da aplicação de uma pesquisa. Para este processo podemos utilizar outra técnica de IA, Agentes Inteligentes (Apêndice 1).

Ainda no programa de elaboração do instrumento de pesquisa, mandaríamos publicar a pesquisa num servidor, o qual teria uma base de dados de pesquisas a serem respondidas. Ao enviar a pesquisa, por meio de um Agente, teríamos um outro programa (Agente) recepcionando, de forma a disponibilizar a pesquisa na Web. O programa instalado no servido (agente) teria algumas funções bem definidas: publicar a pesquisa, verificar o término da pesquisa pela data ou pelo número de pesquisado e devolver ao pesquisador os dados colhidos.

Após a expiração da data ou do número de pesquisados, voltariamos ao programa construtor de instrumentos de pesquisa para então analisar os dados. Estes poderiam ser gravados em arquivos de forma a serem lidos por software estatísticos para a construção de análises, gráficos, etc.

Por outro lado, nem sempre se tem as perguntas certas para a elaboração do instrumento de pesquisa, justamente por não sermos especialista em construir instrumentos de pesquisas. Imaginado que vários usuários teriam o programa de elaboração de pesquisa e que todos a publiquem no mesmo servidor, podemos tirar vantagens desse fato. Por exemplo, gostaríamos de aplicar uma pesquisa sobre um determinado público alvo e, no entanto, já sabemos que pesquisas semelhantes foram aplicadas, assim, através de um Agente Inteligente, utilizando a técnica de Raciocínio Baseado em Casos (exemplares) (Apêndice 2), buscamos sobre o servidor os instrumentos já publicados.

O mesmo ocorreria se necessitássemos de perguntas referentes a um determinado publico alvo. O banco de dados de pergunta de um determinado usuário do programa de elaboração de pesquisa é diferente do banco de dados de pergunta de outro, pois cada um elabora as perguntas conforme seu interesse pessoal, o que nos leva a pensar que não teremos localmente perguntas sobre vários assuntos, mas sim virtualmente. A questão então é enviar um Agente Inteligente, utilizando Raciocínio Baseado em Casos, em todos os usuários possuidores do programa, para buscar as perguntas referentes a um público alvo que se quer montar um instrumento de pesquisa.

A elaboração de um sistema como este, envolve várias tecnologias, o que não o torna nada trivial. O próximo capítulo descreve o sistema desenvolvido bem como as tecnologias envolvidas.

MODELO TEÓRICO DE SISTEMA DE PESQUISA

3.1 Introdução

A Figura 5 mostra o processo de elaboração, aplicação e análise de pesquisas, que até então vinham sendo aplicados.

Os passos descritos podem ser informatizados, utilizando-se das tecnologias disponíveis. O uso destas trazem benefícios aos agentes envolvidos num processo de pesquisa tais como:

- Na confecção da pesquisa – *software* de apoio à elaboração das pesquisas, utilizando agentes inteligentes (RBC) para busca de pesquisas e questões;
- No ajuste da pesquisa – através de técnicas de IA, como lógica *fuzzy*;
- Na aplicação da pesquisa – um software que proporcione periodicidade na aplicação das pesquisas, bem como redução de custos e adestramento, independente da quantidade de pesquisados. Agente que publique e avise o término da pesquisa;
- Na análise das respostas da pesquisa – análise de variáveis relativas a conceitos cognitivos e migração dos dados colhidos a outros *softwares* dedicados a estatística. Análise parcial dos dados através de agrupamentos.

Sistema Tradicional de Pesquisa

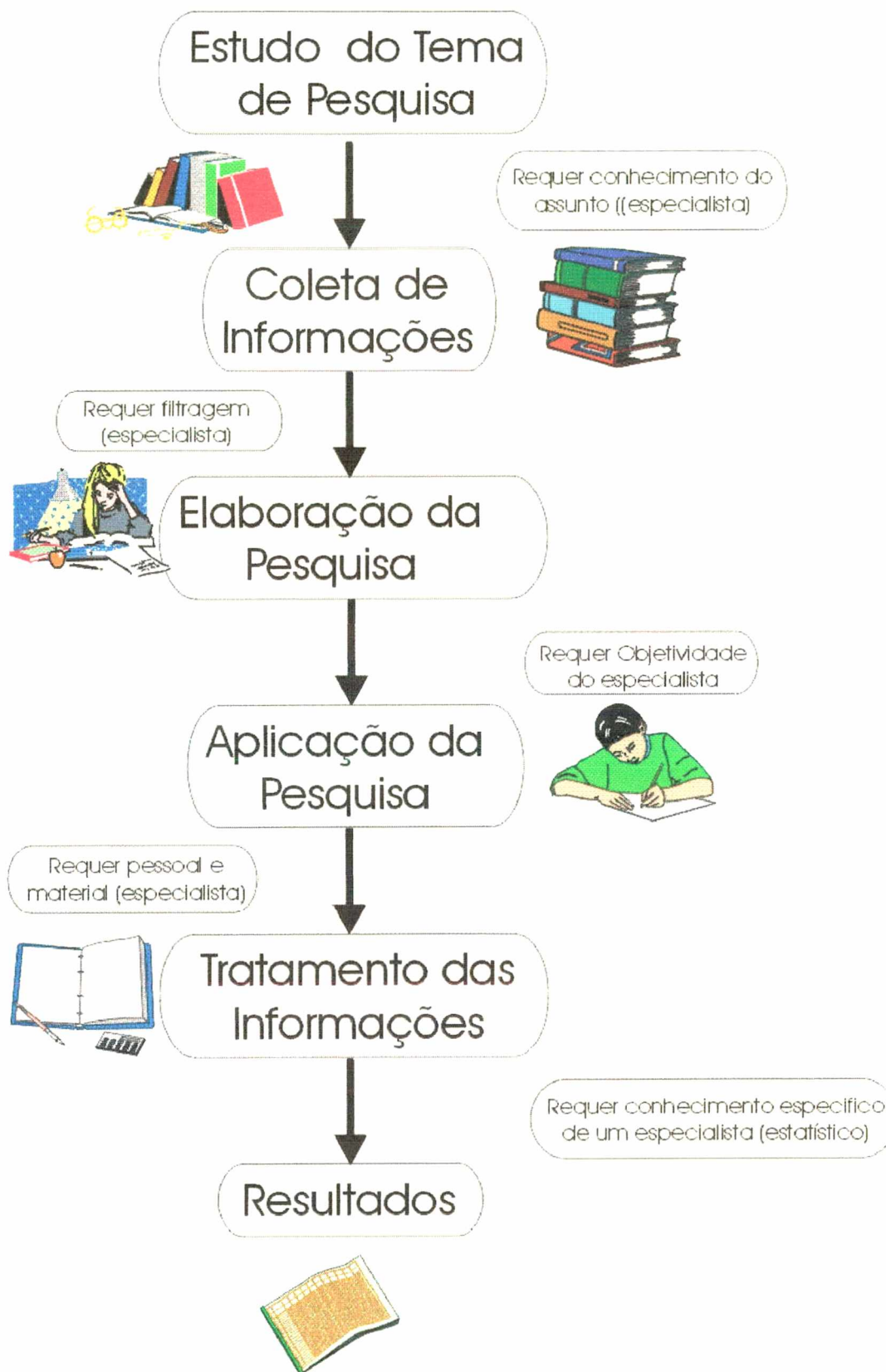


Figura 5. Sistema de pesquisa proposto.

Assim, é proposta uma nova arquitetura para o modelo tradicional onde envolve o ambiente Web, Figura 6.

ARQUITETURA DO MODELO PROPOSTO

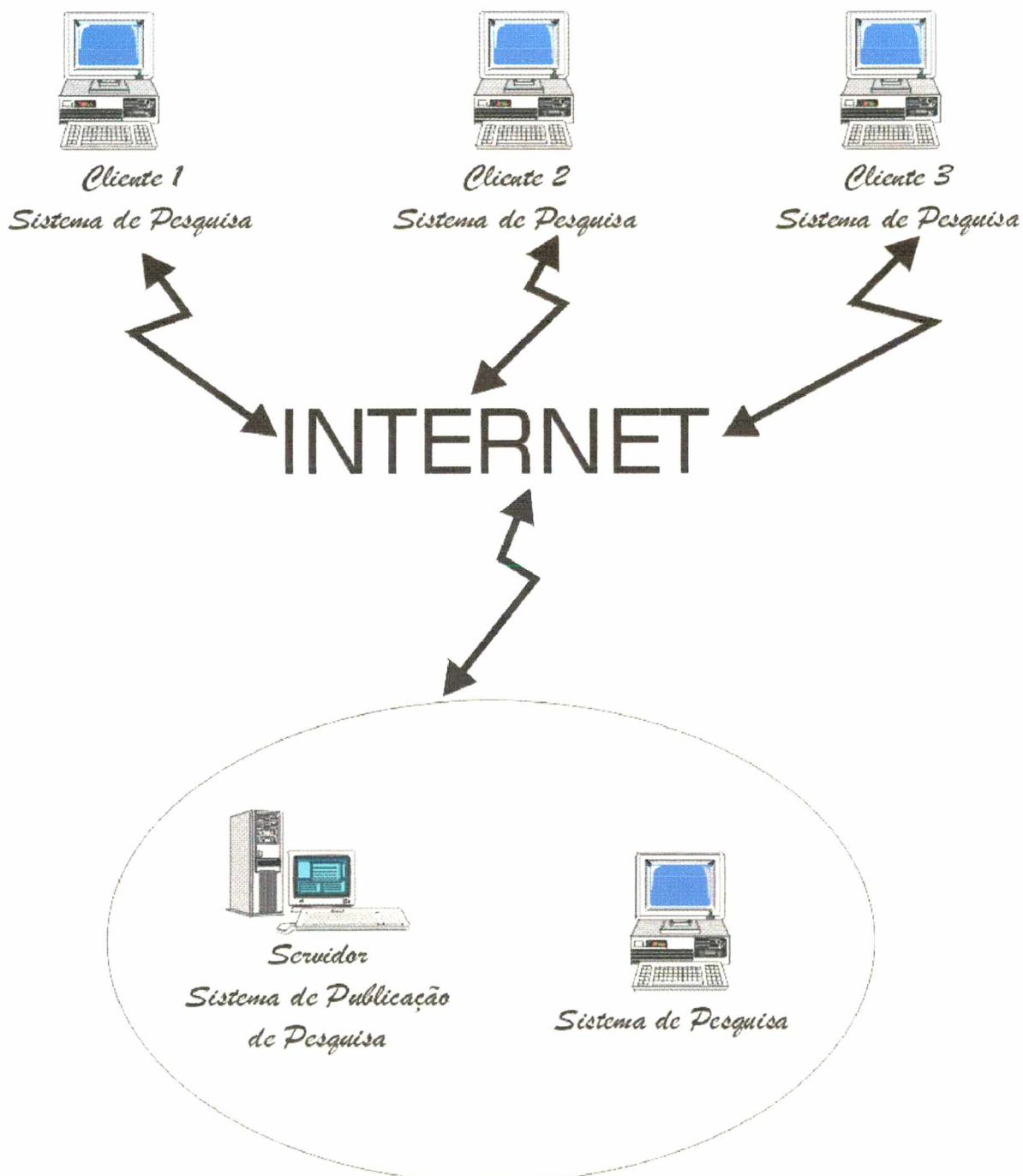


Figura 6. Arquitetura do sistema proposto.

Como se pode observar na Figura 6, existem três módulos distintos:

- Módulo do Cliente: este terá a possibilidade de elaborar a pesquisa utilizando-se do Sistema de Pesquisa, bem como de agentes inteligentes que irão procurar nos demais clientes e servidor, pesquisas e perguntas sobre determinado assunto. Este ainda contém, além de outros, um agente de publicação de pesquisa que ficará responsável pela publicação e término da pesquisa;
- Módulo Servidor: semelhante ao módulo cliente, isto é, com a mesma funcionalidade, além de gerenciamento dos demais clientes e controle de IP. Neste módulo ficam os principais agentes;
- Módulo de Publicação de Pesquisa: é neste módulo que os pesquisados irão responder as pesquisas via *browser*. As pesquisas publicadas estarão aqui para serem respondidas e enviadas ao servidor.

Desta forma, segue detalhadamente cada um destes módulos.

3.2 Implementação do Sistema de Pesquisas

O sistema proposto é um aplicativo desenvolvido à investigação de cenários, diante às necessidades das diversas áreas das ciências sociais.

O sistema em questão trabalhará com questionários. Neste pode-se elaborar perguntas de acordo com o item 5.4.1 e 5.6. Os demais itens são intrínsecos ao agente pesquisador, isto é, há um especialista, e o aplicativo dará suporte como veremos mais adiante.

Este sistema não é somente um aplicativo de elaboração de pesquisa, mas operacionaliza o instrumento sob qualquer plataforma informatizada, oportunizando o pesquisado responde-lo em tempo real ou não, gerando resultados para posterior tratamento analítico.

De acordo com a Figura 7 a operacionalização do sistema está dividida em três momentos distintos: Elaboração, Aplicação e Análise.

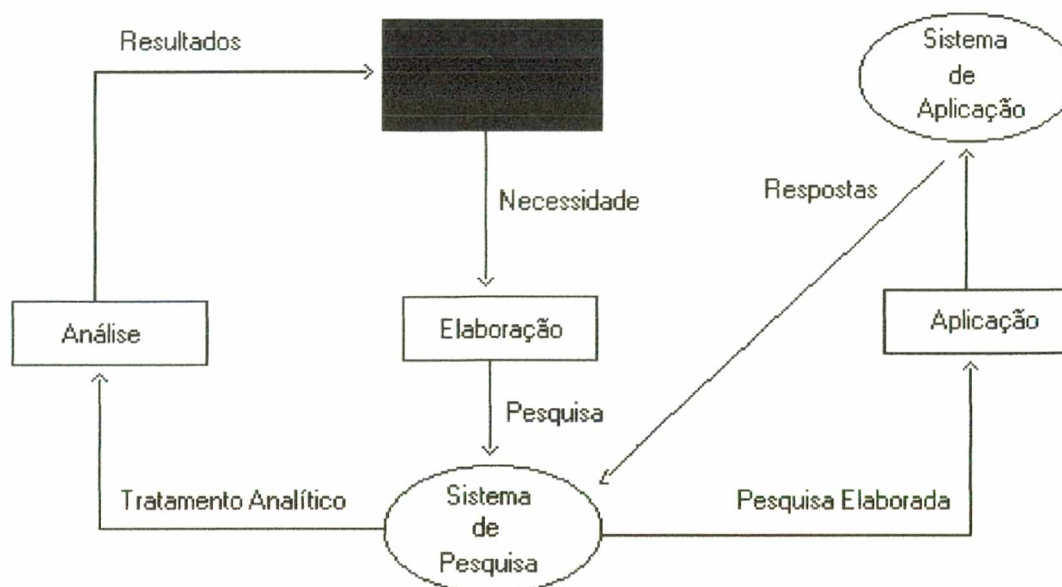


Figura 7. Elaboração, Aplicação e Análise.

Elaboração: dado um problema real, criou-se uma necessidade. O pesquisador vai criar e elaborar um instrumento de acordo com a necessidade requisitada, utilizando o sistema desenvolvido para elaboração de pesquisa.

Aplicação: definido o instrumento, esse passa a ser disponibilizado ao pesquisado por meio do Sistema de Publicação de pesquisas via Internet. Ao pesquisado cabe responder as questões, enquanto estas retornam para o sistema de pesquisa servidor.

Análise: findo a aplicação os dados retornarão e o avaliador verifica os resultados de duas maneiras. A primeira possibilidade de verificar o resultado é dividida em três outras. *Gabarito* é uma análise das respostas, baseado numa amostra respondida pelo pesquisador, de forma a verificar os erros e acertos do pesquisado. O resultado é verificado em forma de porcentagem. *Emia* é uma análise das respostas, baseado na escolha de duas perguntas por parte do pesquisador, de forma a agrupar a massa de resposta pela Ancora das perguntas escolhidas. Para exemplificar, suponha ter as seguintes perguntas, entre outras: 1) Você é? Aluno, Professor ou Funcionário 2) Qual seu partido político ? a, b, c ou d. O sistema agruparia, em forma de porcentagem, todos

os alunos do partido a, depois b, depois c, depois d, e assim para professor e funcionário. *Agrupamento* é uma análise das respostas, baseado na escolha de algumas perguntas, por parte do pesquisador, para que verifique a percentagem de acertos em determinadas perguntas. Este item torna-se interessante pois conseguimos detectar, na amostra de resposta, a percentagem de pessoas que pensam iguais a respeito de um determinado assunto.

A segunda possibilidade é gerar um arquivo com os dados para posterior análise em um software aplicado à estatística.

O sistema apresenta o módulo para a definição dos parâmetros da pesquisa representados por objetos que compõem a pesquisa. São parâmetros da pesquisa: Público Alvo, Critério, Perguntas e Âncoras.

3.2.1 Público Alvo

Público Alvo é o agrupamento de pessoas, ligadas por interesse social, econômico, político, religioso, intelectual, etc., isto é, pessoas a quem se destina a pesquisa a ser elaborada.

Público alvo define a clientela a ser atingida pela pesquisa. Este parâmetro tem relacionamento direto com as perguntas elaboradas pela pesquisa. Para cada pesquisa construída, será necessário defini-lo, a fim de executar o primeiro filtro de perguntas.

O pesquisador ao interagir com o sistema definirá o seu público alvo, o qual pode ou não estar cadastrado. A escolha do público alvo se dá na concepção da pesquisa, que, automaticamente, filtrará as questões concernentes ao novo instrumento. Caso o público alvo desejado não esteja no rol, este deverá inseri-lo.

3.2.2 Critérios

Critérios são usados para distinguir as pesquisas através de predicados relevantes. A utilização deste parâmetro é feita através de pesos, quantificando o grau de importância de cada critério para a pergunta (item 6.3 aborda com mais detalhes).

É através dos critérios que será possível escolher as perguntas que comporão uma determinada pesquisa, baseada no tratamento multicritério difuso.

Assim, esse parâmetro só terá relevância quando da aplicação de seleção das perguntas através do multicritério difuso, pois ele dá um grau de especificidade a pesquisa, agregando os valores do pesquisador.

Os critérios estão intimamente ligados ao público alvo, pois um é subsistema do outro. Um público alvo pode conter n critérios, no entanto, definiu-se que uma pesquisa conterá entre 0 (zero) e 10 (dez) critérios, como descreve Augras (1974)

3.2.3 Perguntas

As perguntas do questionário são os componentes mais importantes de uma pesquisa, pois tem um efeito fundamental nos resultados da pesquisa.

O resultado obtido com o instrumento gerado está intimamente ligado com as questões, pois o superveniente a essas é o objeto fim do instrumento. Quando mal elaborada, os resultados podem ser catastróficos. Assim, o pesquisador deve estar intimamente ligado ao objetivo da pesquisa a fim de que as questões estejam no contexto desejado.

Três aspectos relevantes devem ser observados na elaboração das perguntas: objetividade, simplicidade e brevidade. As perguntas elaboradas podem ser interrogativas ou não. É importante definir o tipo de resposta que se deseja.

Procura-se evitar perguntas que tenham palavras com sentido ambíguo, usem palavras muito enfáticas, não sejam aplicáveis, exijam muito da memória, tenham duas questões implícitas e sejam tendenciosas, como visto no Capítulo V.

Para tanto, o sistema apresenta uma interatividade flexível de pergunta e resposta, a fim de personalizar cada pesquisa.

3.2.4 Âncoras

Âncora é o parâmetro utilizado para definir o formato de escolha associado a uma resposta. Uma pergunta está, via de regra, associada a uma âncora. Sendo que aquela pode ser utilizada por vários tipos de âncoras, em diferentes pesquisas.

Todas as questões assumem a âncora sim/não, por *default*, ao criarmos uma pergunta, no entanto, a qualquer momento pode ser alterada ou elaborada nova âncora.

Ao definir uma nova âncora, deve-se a ter a quantidade de alternativas. É uma escolha do pesquisador, isto é, onde entra o conhecimento do especialista. Ao decidir por uma âncora com cinco alternativas, por exemplo, obtém-se um ponto médio, neutro ou tendencioso. Usando seis alternativas espalha-se na âncora a tendência, principalmente quando a questão é pessoal.

3.3 Métodos de Seleção de Perguntas

O pesquisador, após a definição da pesquisa fará a seleção das perguntas que irão compor o questionário. Para isso, o pesquisador poderá optar pela escolha aleatória de um determinado número de perguntas ou através do método de multicritério difuso.

A Figura 8 mostra o modelo tradicional de escolha de perguntas, isto é, se determinada pergunta participaria ou não do questionário. Também mostra o novo sistema adotado, onde se usa multicritério difuso fazendo com que o pesquisador coloque seu *animus* na escolha das perguntas que farão parte do questionário, reformulando consideravelmente o modo de escolha das perguntas.

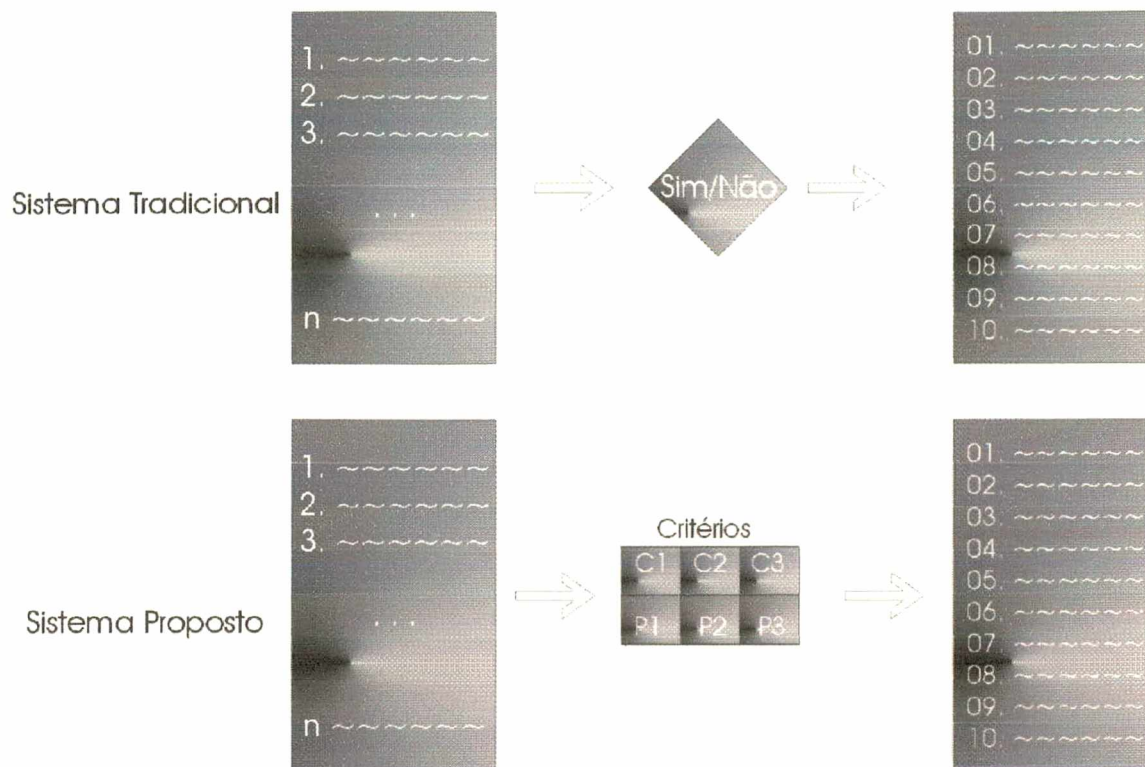


Figura 8. Escolha de perguntas

3.3.1 Método de Multicritério Difuso

Ao utilizar o método de multicritério difuso definirá os critérios que deseja evidenciar na pesquisa.

	C ₁ Pontualidade	C ₂ Metodologia	C ₃ Didática
P ₀	7	10	4
P ₁ Cumpriu a ementa?	4	6	0
P ₂ Mostrou Domínio do assunto?	0	3	6
P ₃ Usou interativismo e dinamismo nas aulas?	0	4	6
P ₄ Foi pontual?	10	0	0
P ₅ Indicou fontes de pesquisas?	7	10	4

Tabela 3. Exemplo de uma pesquisa.

Primeiramente, o Agente Pesquisador definirá o peso que quer dar a cada critério, a este chamaremos de perfil da pesquisa P_0 . Em seguida, para as perguntas já filtradas pelo público alvo, estabelecerá o peso que cada pergunta representa para os critérios definidos.

A Tabela 3 apresenta um exemplo de uma pesquisa a ser aplicada em uma sala de aula, onde o público alvo escolhido foi o professor com os seguintes critérios: pontualidade, metodologia, didática.

Se os valores fossem *crisp* a pergunta j para o critério i : C_i^j seria 0 ou 1. Como o tratamento é difuso, 3 é o peso dado a pergunta P_2 , considerando o critério C_2^2 . Através deste processo de escolha de perguntas o pesquisador estará colocando seu *animus*, o que torna o sistema democrático.

3.3.1.1 Resolução do Problema

$$q^j = 1 - \frac{\sum_{i=1}^m |\mu_i^0 - \mu_i^j|}{m + \sum_{i=1}^m \mu_i^0 \wedge \mu_i^j}$$

Onde: q^j é o módulo da distância entre o perfil P_0 e outro; m é o número de critérios.

Van Velthoven (1975) apud Kickert (1978) sugere, para uso desta equação os seguintes passos:

- Definir os conjuntos suporte;
- Definir a função de pertinência requerida para o perfil P_0 ;
- Encontrar a função de pertinência para as demais perguntas $P_1 \dots P_5$;
- Computar a seleção.

Destarte, primeiramente definiu-se o conjunto suporte pergunta, Figura 8: Fraca, Média, Forte. A definição desta variável linguística deu-se pelo fato de não

termos para cada critério um conjunto suporte diferente, e consideramos para todos os critérios o mesmo conjunto, baseado no conjunto suporte pergunta.

Como descrito anteriormente, item 4.2.2, cada pesquisa pode definir de 0 (zero) a 10 (dez) critérios, o que tornaria difícil definir funções de pertinência para cada critério, mesmo porque estes são adjetivos, e representam cognitivamente fatos diferentes para cada indivíduo, daí definir o conjunto suporte para pergunta Figura 9.

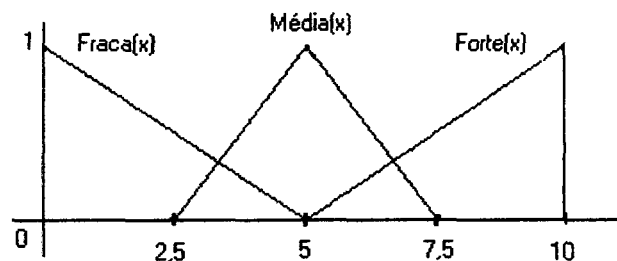


Figura 9. Conjunto suporte para pergunta.

O Agente Pesquisador deverá estabelecer pesos para cada pergunta/critério dentro do intervalo de 0..10 e estes serão fuzificados para cada critério de acordo com as suas funções de pertinência:

$$Fraca(x) = \begin{cases} 0 & x > 5 \\ 1 - \frac{x}{5} & x \leq 5 \end{cases}$$

$$Média(x) = \begin{cases} 0 & x \leq 2,5 \\ 1 - \left(\frac{5-x}{2,5}\right) & 2,5 < x \leq 5 \\ \frac{7,5-x}{2,5} & 5 < x \leq 7,5 \\ 0 & x > 7,5 \end{cases}$$

$$Forte(x) = \begin{cases} 0 & x < 5 \\ 1 - \left(\frac{10-x}{5}\right) & x \geq 5 \end{cases}$$

Em seguida foi definido o conjunto pertinência para o perfil P_0 . A Tabela 4 mostra também os demais conjuntos pertinência para cada pergunta.

	C ₁	C ₂	C ₃	Fraca	Média	Forte	Fraca	Média	Forte	Fraca	Média	Forte
P ₀	7	10	4	0,0	0,2	0,4	0,0	0,0	1,0	0,2	0,6	0,0
P ₁	4	6	0	0,2	0,6	0,0	0,0	0,6	0,2	1,0	0,0	0,0
P ₂	0	3	6	1,0	0,0	0,0	0,4	0,2	0,0	0,0	0,6	0,2
P ₃	0	4	6	1,0	0,0	0,0	0,2	0,6	0,0	0,0	0,6	0,2
P ₄	10	0	0	0,0	0,0	1,0	1,0	0,0	0,0	1,0	0,0	0,0
P ₅	0	0	0	0,0	0,2	0,4	0,0	0,0	1,0	0,2	0,6	0,0

Tabela 4. Conjunto pertinência para cada pergunta.

Assim, teremos o conjunto pertinência requerido para o perfil P₀ como sendo:

$$P_0 = \{Fr/0,0; Me/0,2; Fo/0,4; Fr/0,0; Me/0,0; Fo/1,0; Fr/0,2; Me/0,6; Fo/0,0\}$$

Para as demais perguntas, P₁ .. P₅, segue conforme Tabela 4.

O último passo proposto por Van Velthoven será o cômputo, a fim de encontrar os melhores *ranks*.

A aplicação direta da fórmula de Índice de Similaridade não foi possível por dois motivos:

- A fórmula aplica-se uma dada pertinência e não um conjunto de pertinência;
- O valor q^j deve satisfazer a seguinte condição: $0 \leq q^j \leq 1$.

Então se adaptou a fórmula:

$$q^j = 1 - \frac{\sum_{i=1}^m \left(\frac{\sum_{k=1}^s |\mu_{i,k}^0 - \mu_{i,k}^j|}{s} \right)}{m + \sum_{i=1}^m \left(\frac{\sum_{k=1}^s |\mu_{i,k}^0 \wedge \mu_{i,k}^j|}{s} \right)}$$

Onde:

q^j é o módulo da distância entre o perfil P₀ e outro; m é o número de critérios.

m é número de critérios

s é o número de variáveis linguísticas num conjunto difuso

Ao aplicarmos a fórmula, como mostra a Tabela 5, teremos o *rank* das perguntas que farão parte da pesquisa.

Perguntas	Rank
Cumpriu a ementa?	0,62
Mostrou domínio do assunto?	0,60
Usou interativismo e dinamismo nas aulas?	0,56
Foi pontual?	0,57
Indicou fontes de pesquisas?	1,00

Tabela 5. Rank das Perguntas.

A fórmula nos dá a distância entre o perfil idealizado e as perguntas sugeridas, então a menor distância será a melhor pergunta e a classificação é verificada em ordem crescente, mas na fórmula proposta tem o valor 1 (um) diminuído do restante o que nos dá a classificação direta da melhor para a pior pergunta (decrecente).

Conforme mostra a Tabela 5, as perguntas classificam-se em ordem decrescente: P₅, P₁, P₂, P₄ e P₃, da melhor para a pior pergunta.

3.3.2 Método Aleatório

A aplicação do método aleatório é uma segunda maneira de selecionar perguntas, auxiliando o pesquisador numa tarefa árdua.

Na sua maioria, as linguagens de computador vêm com algumas funções e/ou procedimentos que implementam números aleatórios. No entanto, nem sempre condiz com nossa realidade, o que nos leva a implementar outros métodos.

Cabe ressaltar que os números aleatórios gerados por computador são pseudo-aleatórios, já que o computador é essencialmente determinístico em sua programação, pois não podem gerar valores verdadeiramente aleatórios, o que vem mais uma vez reforçar o método descrito na seção anterior.

Os problemas encontrados com a geração de números aleatórios estão calçados na semente que gerarão os demais números. Assim, os métodos preocupam-se em gerar

uma semente que não venha resultar números repetidos. Desta forma, destacamos o Método de Congruência Linear.

Este método parte de um valor inicial, semente, e a cada chamada da função, usa os últimos valores gerados para gerar os próximos. Se r_i é um resultado gerado na chamada i , então o próximo valor será gerado:

$$r_{i+1} = (ar_i + b) \text{Mod } m$$

Onde, a , b e m são valores fixos.

Para exemplificar, seja $a=13$, $b=3$ e $m=17$, com $r_0=1$ e 4 iterações, como mostra a Tabela 6.

Valor	Cálculo	Valor Final
R0		1
R1	$(13*1+3) \text{ mod } 17$	16
R2	$(13*16+3) \text{ mod } 17$	7
R3	$(13*7+3) \text{ mod } 17$	9
R4	$(13*9+3) \text{ mod } 17$	1

Tabela 6. Exemplo de iterações.

Para seleção das perguntas, usou-se o resultado deste método, associando cada valor a uma pergunta, e se resultar em número par esta é uma pergunta a ser selecionada. O pesquisador define o número de perguntas a participar da pesquisa, de forma que as n perguntas com maior *rank* comporão a pesquisa, Tabela 7.

Pergunta	Valor
Cumpriu a ementa ?	16
Mostrou Domínio do assunto?	7
Usou interativismo e dinamismo nas aulas?	9
Foi pontual?	1

Tabela 7. Rank das perguntas.

3.4 Arquitetura de Agentes Para o Sistema Proposto

A arquitetura proposta é uma arquitetura básica composta por vários agentes que se comunicam entre si a fim de alcançarem os resultados no processo de elaboração de pesquisa.

Esta é composta de duas classes de agentes, como mostra a Figura 10, Agentes da Classe Pesquisador e Agentes da Classe Pesquisas.

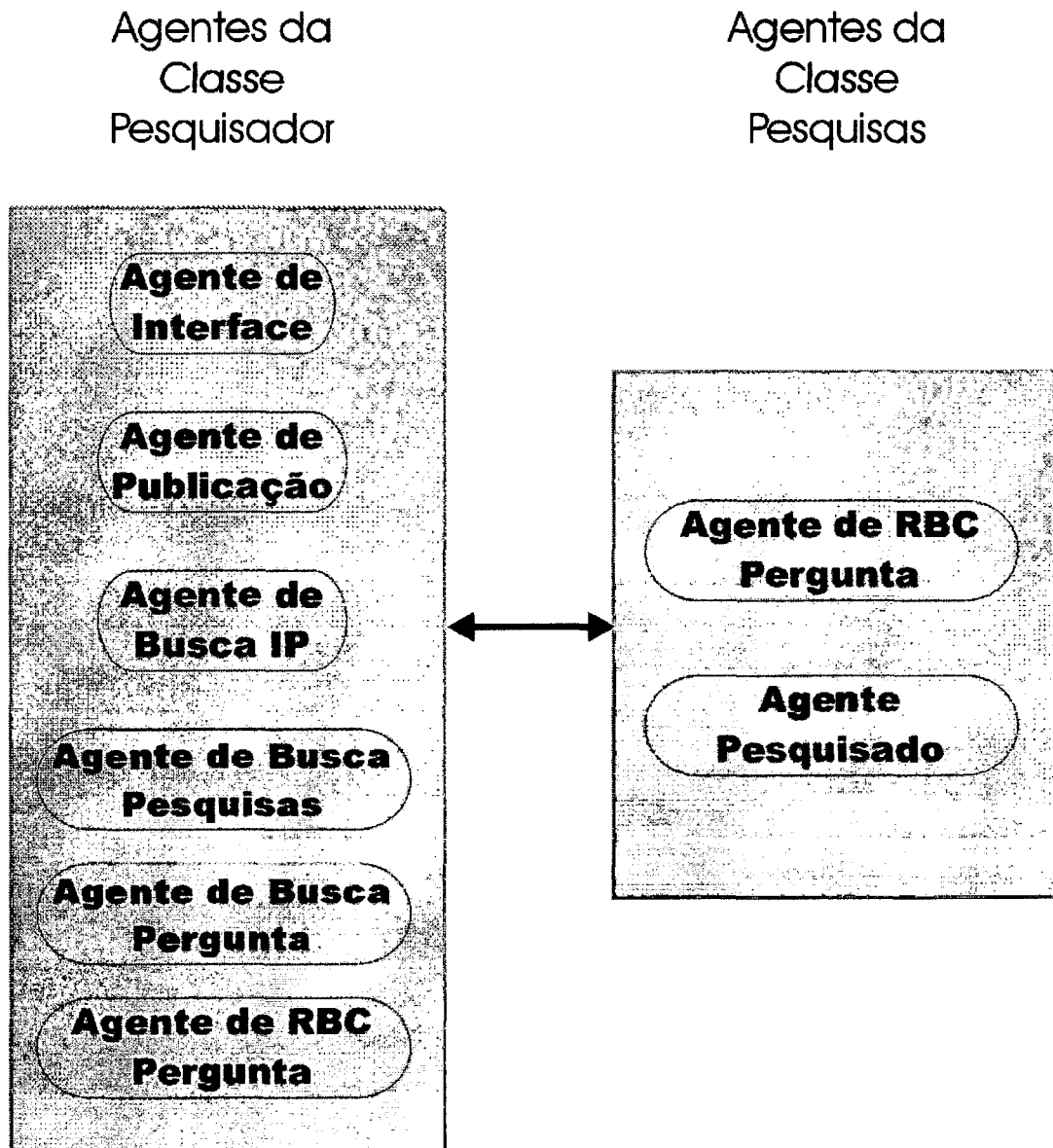


Figura 10. Classes de agentes.

Ao analisar a Figura 10 deve-se ter em mente que há várias Classes de Pesquisador trocando informações entre si e/ou com os Agentes da Classe Pesquisas.

Como visto no item 1.6, uma arquitetura de agentes estabelece normas de comunicação entre agentes, exerce o agendamento de eventos, armazena a base de conhecimento, aciona a máquina de inferência e implementa o protocolo de comunicação.

Sabe-se que existe no mercado plataformas de desenvolvimento e execução de agentes, como o Sistema Concordia, Agent TCL, Telescript, TACOMA, Mole, Aglet WorkBench etc. e cada qual com sua linguagem própria de comunicação, pois não há um consenso, um padrão para comunicação de agentes. Como exemplos de linguagens de comunicação de agentes pode-se citar Agent Talk KIF, KQML, ACL etc.

Além das plataformas e linguagem de comunicação a escolha de uma linguagem de programação deve ser estabelecida e, dentre as mais usadas, cita-se Java, LALO, Obliq, Phantom, Tcl/Tk, Telescript.

Assim, neste sistema optou-se por não usar uma plataforma de desenvolvimento nem uma linguagem de comunicação, das descritas acima, de forma que os agentes relacionados comunicam-se através de *brokers*. Como linguagem de programação escolheu-se o Delphi, por estar todo o sistema inserido na mesma. Já a parte de publicação de pesquisa utilizou-se Java Applet.

3.4.1 Agentes da Classe Pesquisador

Os agentes descritos nesta seção são responsáveis por amenizar o trabalho do pesquisador na busca de informações, elaboração de pesquisas, elaboração de perguntas, filtragem das perguntas, aplicação da pesquisa e avaliação dos dados.

3.4.1.1 Agente de Interface

Além de manter os dados necessários para realização de pesquisas, serve de interface entre o pesquisador e outros agentes, este tem como finalidades:

- Comunicar o pesquisador que sua pesquisa foi respondida;
- Receber as respostas das buscas de pesquisas e perguntas.

Caracteriza-se por ser um agente estático e por possuir os atributos de autonomia, comunicabilidade, confiabilidade, degradação gradual, reatividade e representabilidade.

3.4.1.2 Agente de Publicação

Findo a elaboração da pesquisa, este é responsável pela publicação da pesquisa na WEB, a fim de que o módulo Publicação de Pesquisa possa habilitá-la a ser respondida. Este também habilita o Agente Pesquisado, o qual é o responsável por detectar o término de uma pesquisa.

Caracteriza-se por ser um agente móvel e apresentar os atributos autonomia, confiabilidade, degradação gradual, reatividade, representabilidade.

3.4.1.3 Agente Busca IP

Este agente está relacionado diretamente ao Agente Busca Perguntas. Quando o Agente Busca Perguntas é acionado, deve ter conhecimento de quais máquinas estão ativas, assim o Agente Busca IP capta esta informação.

Caracteriza-se por ser um agente móvel e apresentar os atributos de autonomia, comunicabilidade, cooperatividade, degradação gradual, responsabilidade.

3.4.1.4 Agente Busca Pesquisa

É responsável por comunicar-se com o Agente RBC Pesquisa e trazer as pesquisas por este encontradas.

Caracteriza-se por ser um agente móvel e apresentar os atributos de autonomia, comunicabilidade, confiabilidade, cooperatividade, degradação gradual, discurso, representabilidade, responsabilidade.

3.4.1.5 Agente Busca Pergunta

Como mencionado no item 6.4.1.3, este ao ser ativado por um pesquisador, comunica-se com o Agente Busca IP, conhecendo quais as máquinas onde este pode buscar perguntas a fim de fortalecer o banco de perguntas sobre determinado tema.

De posse deste conhecimento, interage com o Agente RBC Pergunta de cada máquina, trazendo as perguntas selecionadas.

Caracteriza-se por ser um agente móvel e apresentar os atributos de autonomia, comunicabilidade, confiabilidade, cooperatividade, degradação gradual, discurso, representabilidade, responsabilidade.

3.4.1.6 Agente RBC Pergunta

No item 1.3 definimos agentes e constatamos que autores divergem do conceito de agentes inteligentes, pois definem de acordo com alguns atributos que o mesmo possa ter.

Para efeito deste trabalho, o atributo inteligência é empregado somente para os que agregam técnicas de IA, por compartilharmos com o conceito de Maes, onde define que agentes são programas de computador que aplicam técnicas de inteligência artificial para auxiliar usuários nas tarefas baseadas por computador. Desta forma descartamos, qualquer possibilidade de um agente ser inteligente sem ater-se a esse conceito.

Destarte, o agente em questão recebe informações (limiar de similaridade dado em porcentagem) do Agente Busca Pergunta, e faz a inferência na base de casos.

O cálculo da similaridade consiste da contagem dos critérios do caso armazenado na base que coincidem com os critérios do caso de entrada. A cada critério i do caso armazenado na base igual ao critério i do caso de entrada, soma-se 1 ao número de critérios coincidentes entre os dois casos.

O grau de similaridade entre o novo caso e o caso armazenado na base é dado pelo percentual do número de critérios coincidentes entre os dois casos em função do número de critérios considerados para a comparação destes.

Para este agente foi implementando as etapas do RBC, vistas no item 3.4, com exceção da revisão e da aprendizagem, as quais não cabem ao sistema, uma vez que este se classifica como um Sistema de Raciocínio Baseado em Exemplos. A aprendizagem dá-se de duas maneiras, pela elaboração de novas perguntas ou pela incorporação da resposta do Agente Busca Pergunta a base de dados.

O Agente RBC Pergunta caracteriza-se por ser um agente estático e apresentar os atributos de inteligência, autonomia, comunicabilidade, confiabilidade, cooperatividade, degradação gradual, representabilidade, responsabilidade.

3.4.2 Agentes da Classe Pesquisas

Da mesma forma que os agentes descritos na Classe Pesquisador, os agentes descritos nesta seção, são responsáveis por amenizar o trabalho do pesquisador na busca de informações, elaboração de pesquisas, elaboração de perguntas, filtragem das perguntas, aplicação da pesquisa e avaliação dos dados.

3.4.2.1 Agente RBC Pesquisa

Este agente trabalha, similarmente, ao Agente RBC Pergunta. O Agente RBC Pesquisa recebe informações (limiar de similaridade dado em porcentagem) do Agente Busca Pesquisa, e faz a inferência na base de casos.

O cálculo da similaridade consiste da contagem dos critérios do caso armazenado na base que coincidem com os critérios do caso de entrada. A cada critério i do caso armazenado na base, igual ao critério i do caso de entrada, soma-se 1 ao número de critérios coincidentes entre os dois casos.

O grau de similaridade entre o novo caso e o caso armazenado na base é dado pelo percentual do número de atributos coincidentes entre os dois casos em função do número de atributos considerados para a comparação destes.

Para este agente foi implementando as etapas do RBC, vistas no item 3.4, com exceção da revisão e da aprendizagem, as quais não cabem ao sistema, uma vez que este

se classifica como um Sistema de Raciocínio Baseado em Exemplos. A aprendizagem dá-se de duas maneiras, pela elaboração de nova pesquisa ou pela incorporação da resposta do Agente Busca Pesquisa a base de dados.

O Agente RBC Pesquisa caracteriza-se por ser um agente estático e apresentar os atributos de inteligência, autonomia, comunicabilidade, confiabilidade, cooperatividade, degradação gradual, representabilidade, responsabilidade.

3.4.2.2 Agente Pesquisado

Este agente é ativado quando recebe mensagem do Agente de Publicação. Após a pesquisa ser publicada, o Agente Pesquisado verifica o número de pesquisados da pesquisa em questão e se satisfeito comunica o pesquisador do término da pesquisa.

Caracteriza-se por ser um agente móvel e apresentar os atributos de autonomia, comunicabilidade, confiabilidade, cooperatividade, degradação gradual, discurso, flexibilidade, persistência, pró-atividade, representabilidade, responsabilidade.

CAPÍTULO IV

IMPLEMENTAÇÃO DO SISTEMA PROPOSTO

4.1 Tecnologia Empregada

Para implementação do sistema foi utilizada a linguagem Delphi e Java. O desenvolvimento do sistema de elaboração de pesquisa e o sistema de controle de publicação foram desenvolvidos em Delphi, enquanto que o sistema de publicação de pesquisa na Web utilizou Java Aplet.

Como sistema operacional, utilizou-se o Windows NT e sob este o servidor IIS (Internet Information Server) para suporte a internet.

Como banco de dados a escolha foi pelo Interbase, tanto no servidor quanto no cliente.

Todos os agentes foram desenvolvidos em Delphi, utilizando a tecnologia socket.

4.2 Sistema de Pesquisa e Publicação

A Figura 11 apresenta a tela principal do sistema. Nela encontramos as opções de cadastro do Público Alvo, Critério, Pergunta e Ancora. Dados que estarão no banco de dados para a construção do instrumento de pesquisa.

O sistema é simples de ser operado. Preenche-se a identificação da pesquisa, a validade que terá, isto é o tempo que ficará publicado, o tipo de acesso a pesquisa, se esta vai ou não exigir identificação do pesquisado. Qual o tipo de Ancora será usada nas perguntas.

Metologia de Pesquisa

Público Critério Pergunta Âncora Freq. Correl. Texto Imprime Pesquisas Perguntas Publicação Respostas

Pesquisa 102 Pesquisa CEJURPS

Pesquisa Perguntas da Pesquisa Resultados

Identificação

Descrição Pesquisa CEJURPS

Público Alvo Alunos

Validade

Inicial 15/07/00 Final 15/12/00

Acesso

Senha

Identificação do Entrevistado

Sim

Não

Opcional

Perguntas

Tipo de Resposta

Aberta

Fechada

Mista

Todos Acima

Algoritmo de Seleção

Lógica Difusa

Aleatório

Propõe Critérios

Critérios de Seleção

DESCRICAO	PESO
Satisfação	1
Atendimento	1
Equipamentos	1

Figura 11. Tela principal do sistema.

Por fim, escolhe-se o tipo de algoritmo que selecionará as perguntas que participarão do instrumento de pesquisa. Ao se escolher o Público Alvo, restringe-se a quantidade de perguntas, pois nem todas as perguntas que estão no banco servirão ao Público Alvo em questão. O mesmo acontece, quando se seleciona o tipo de Âncora, pois nem todas as perguntas, por exemplo, serão abertas. Mesmo usando estes filtros, dependendo da quantidade de perguntas cadastradas no banco de perguntas ainda poderão restar muitas perguntas. Assim, na escolha das perguntas restantes, define-se qual algoritmo, Aleatório ou Multicritério Difuso (AMD), para auxiliar o pesquisador. Se a escolha se der pelo aleatório não há necessidade de definir os critérios, mas se a escolha for pelo AMD, define-se os critérios que se quer avaliar em relação ao Público Alvo, passando para a próxima palheta, como mostra a Figura 12.

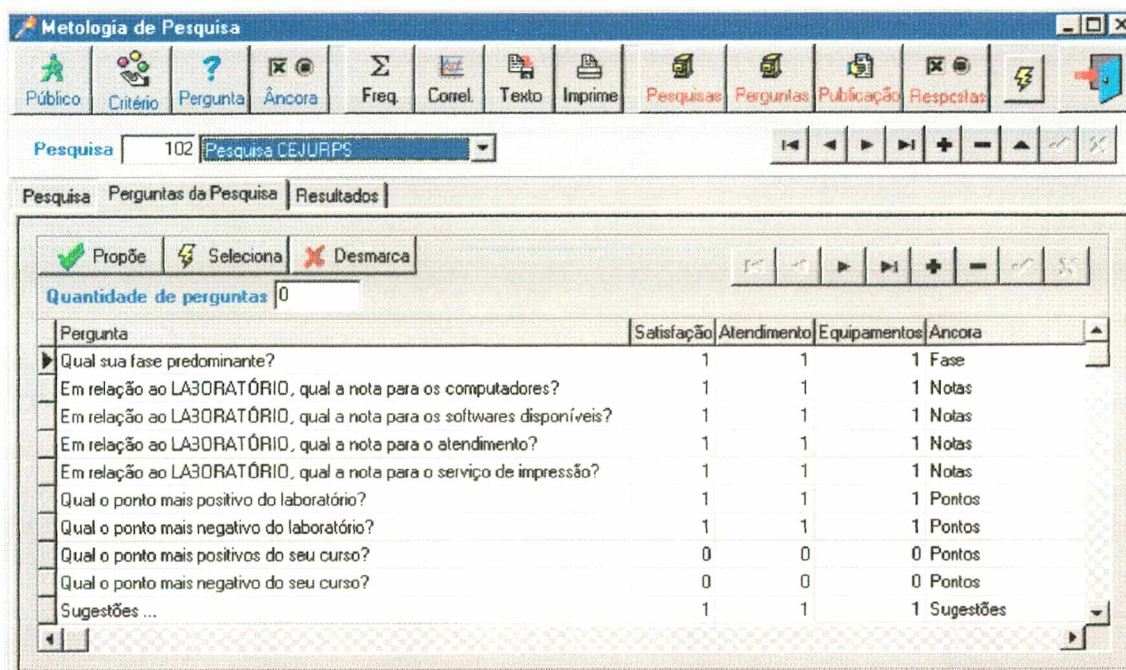


Figura12. Tela perguntas da pesquisa.

Ao clicar no botão propõe, as perguntas filtradas pela tela anterior aparecerão. Se a quantidade de perguntas não for a desejada, define-se então a quantidade, e para cada critério dá-se um peso em relação a pergunta. Ao clicar em selecionar, termina o processo de construção do instrumento, passando-se para a publicação da pesquisa como mostra a Figura 13.

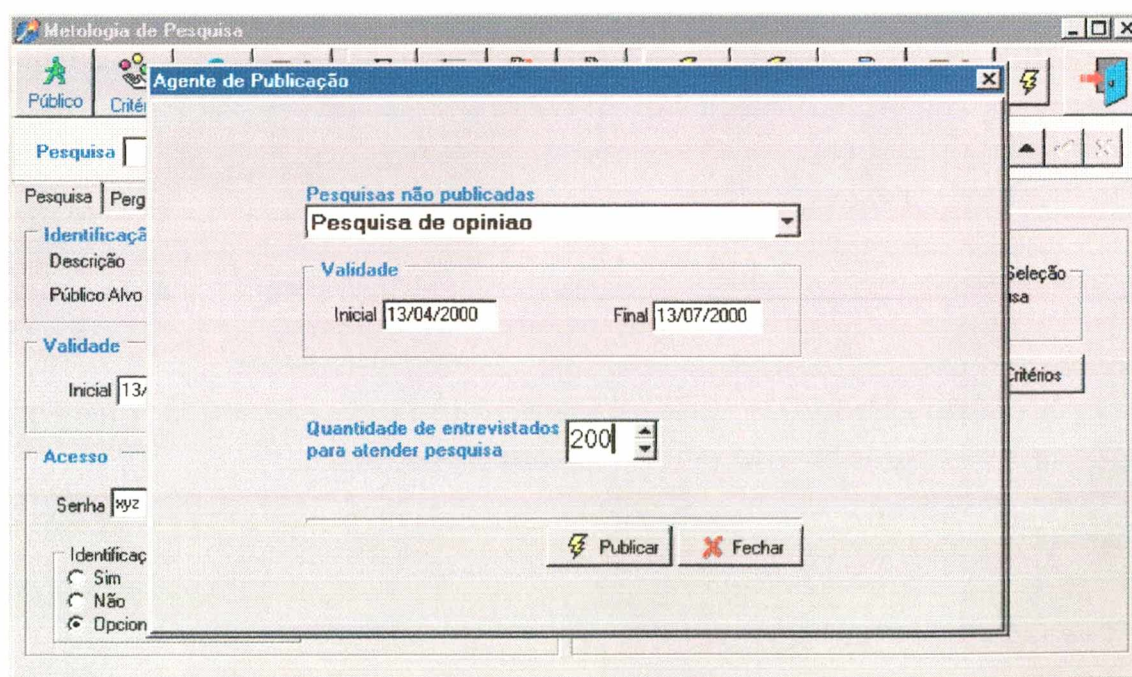


Figura 13. Tela de publicação da pesquisa.

Na tela de publicação, escolhe-se a pesquisa a ser publica, o tempo em que esta permanecerá na Web e o número de pesquisado. Quando expirar a data ou o número de pesquisado for satisfeito, a agente publicação, Figura 14, enviará ao dono da pesquisa os dados da pesquisa em questão, de forma que o pesquisador toma ciência do término da pesquisa.

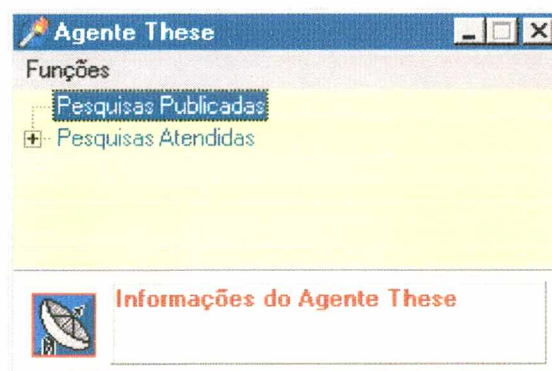


Figura 14. Tela agente.

4.3 Analise dos Dados

Após a agente publicação ter devolvido os dados dos pesquisados tem-se duas possibilidades: salvar a massa de dados em um arquivo .txt, Figura 14 para posterior analise em um software estatístico ou analisar os resultados sob forma de percentagem Figura 15.

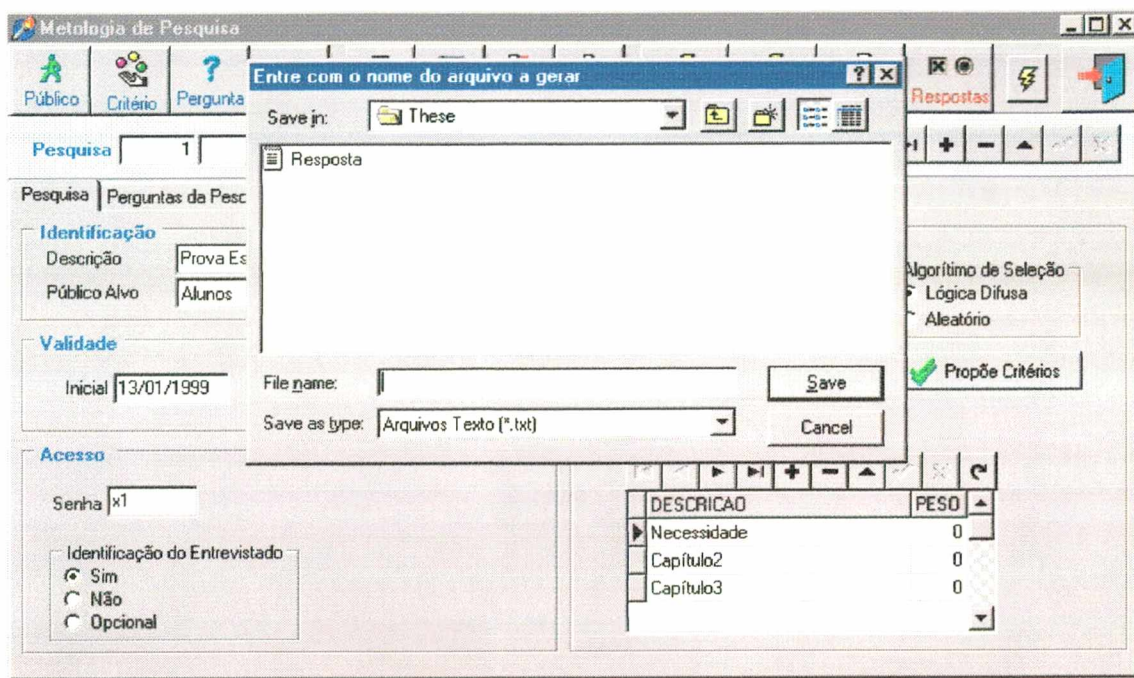


Figura 15. Tela salvar dados da pesquisa.

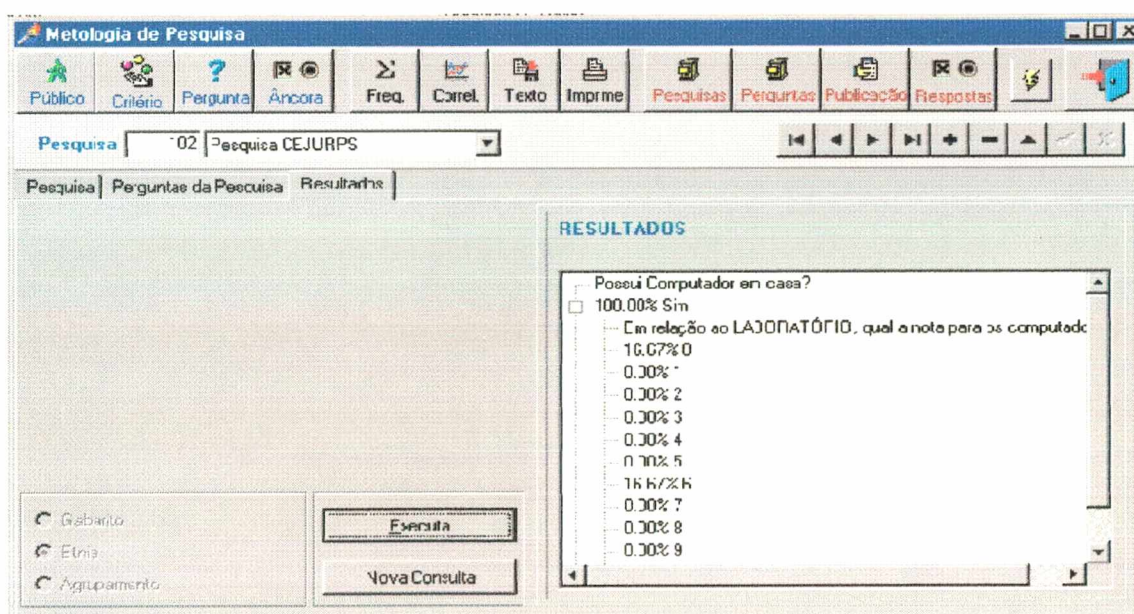


Figura 15. Tela resultado em percentagem.

4.4 Agentes de Busca

Há dois agentes de Busca, um para pesquisas, Figura 16 e outro para perguntas, Figura 17.

Ambos os dois são agentes empregam técnicas de Inteligência Artificial, Raciocínio Baseado em Caso.

O Agente Pesquisa busca no servidor, pois é lá que estão todas as pesquisas publicadas, pesquisas com um grau de similaridade de acordo com o Público Alvo e Critério. O Agente faz o mesmo processo que o Agente de Pesquisa, com a diferença que este busca em todos os clientes perguntas com determinado grau de similaridade.

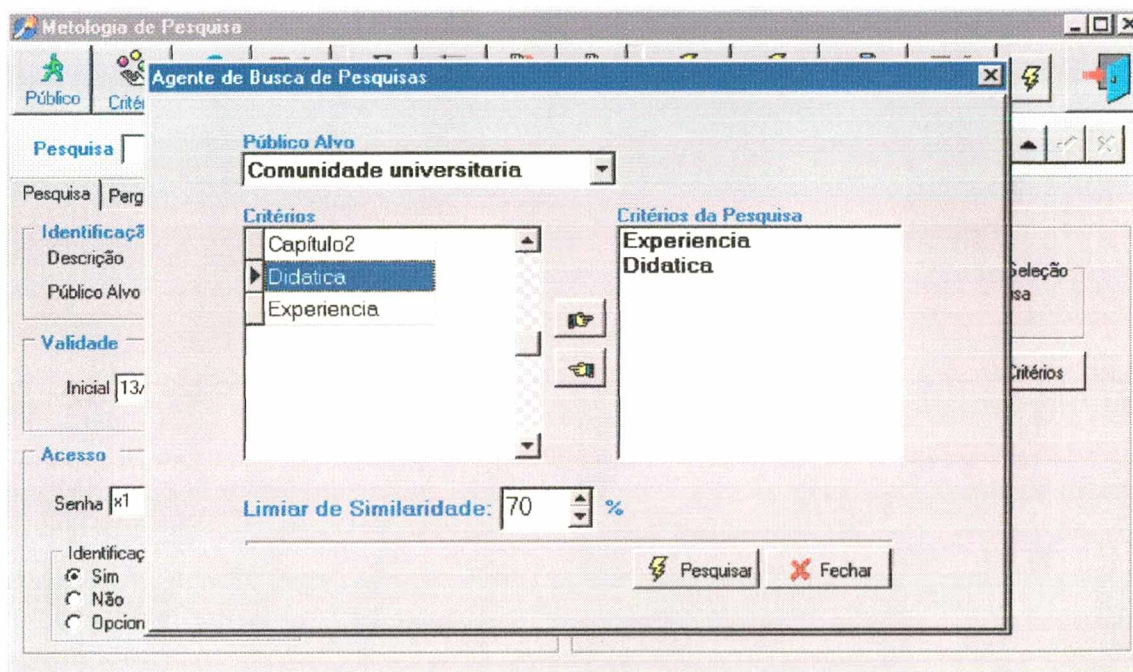


Figura 16. Tela agente de busca pesquisa.

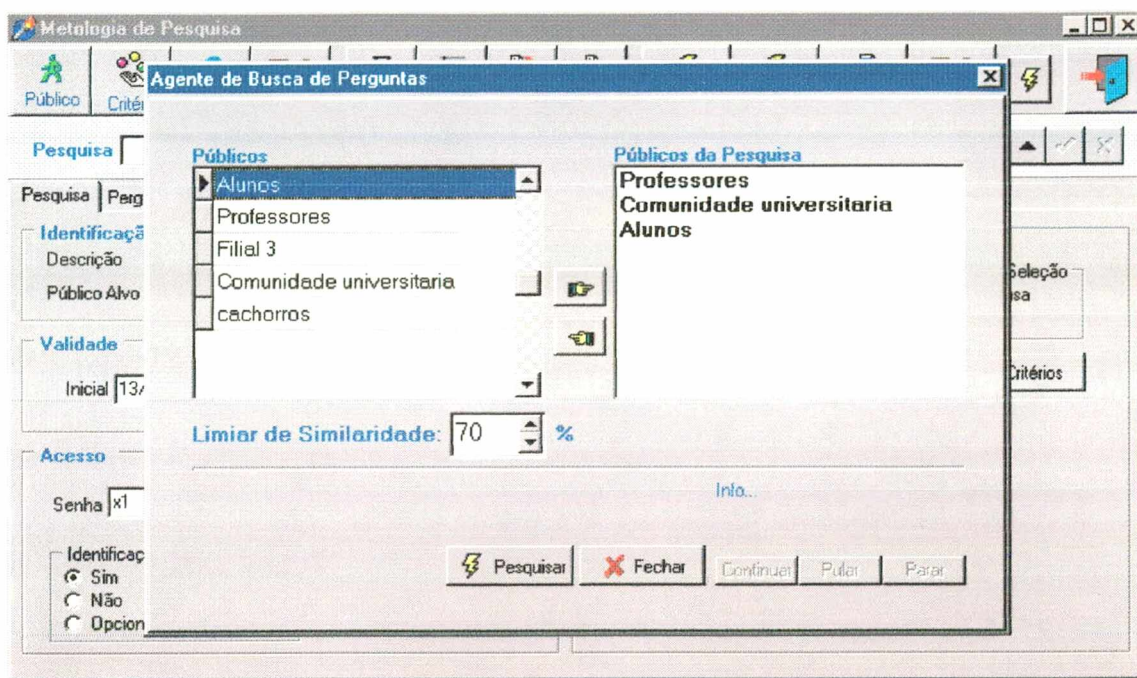


Figura 17. Tela de agente de busca pergunta.

4.5 Sistema na Web

Após ter publicado a pesquisa, Figura 12, a pesquisa estará disponível na web para ser respondida, Figura 18. Neste sistema o pesquisado escolhe a pesquisa a ser respondida, passando a responde-la. Após o pesquisado ter respondido a pesquisa os dados retornarão para o pesquisador através do agente de publicação, Figuras 18, 19 e 20.

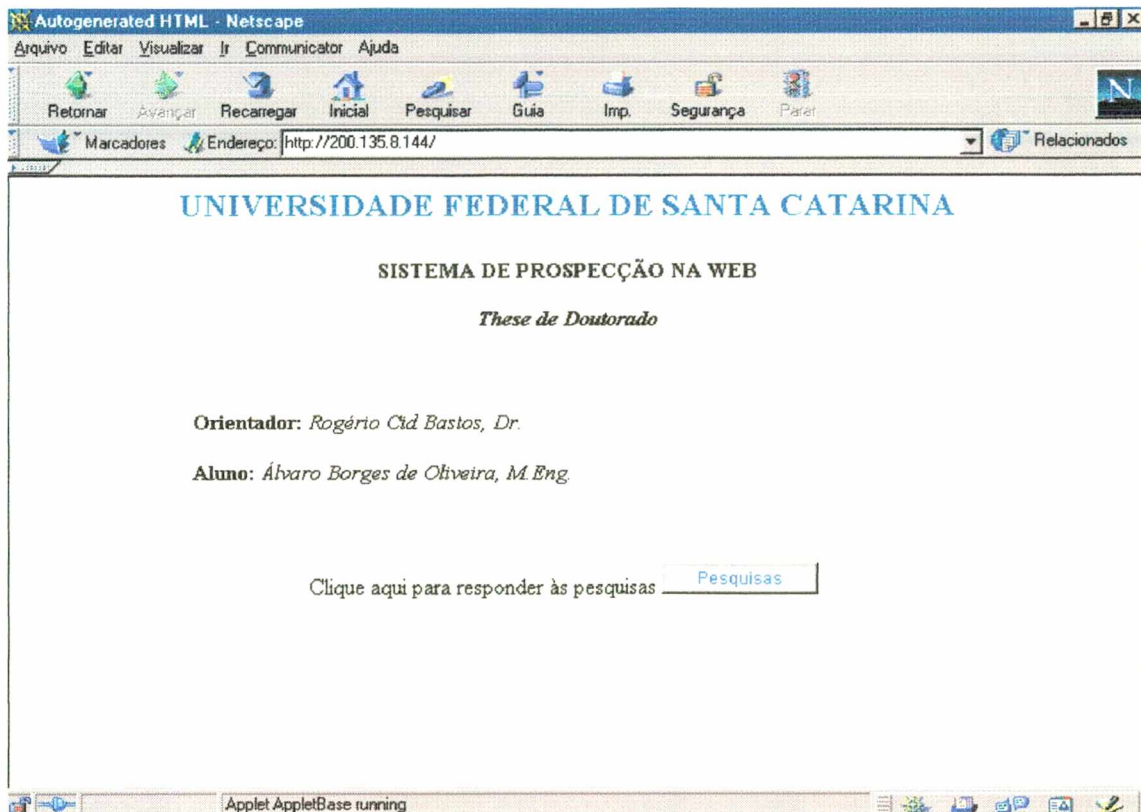


Figura 18. Tela Inicial na Internet.

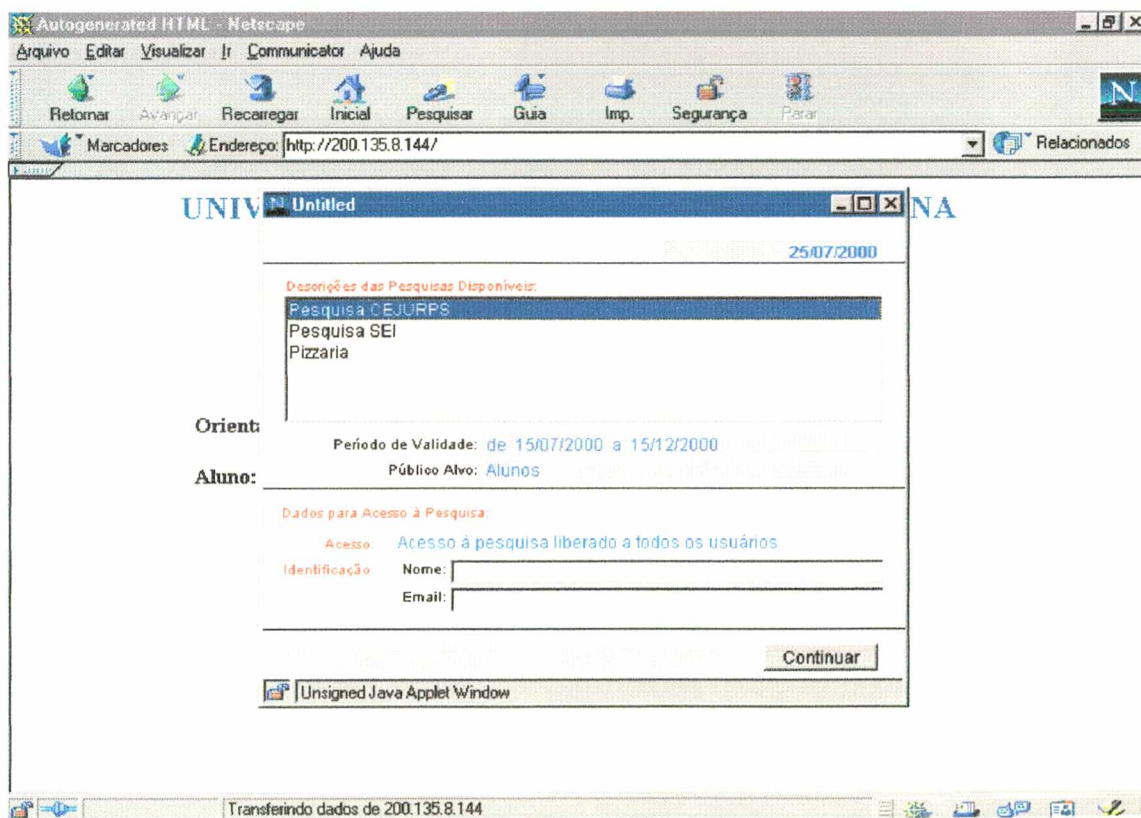


Figura 19. Escolha da pesquisa a ser respondida.

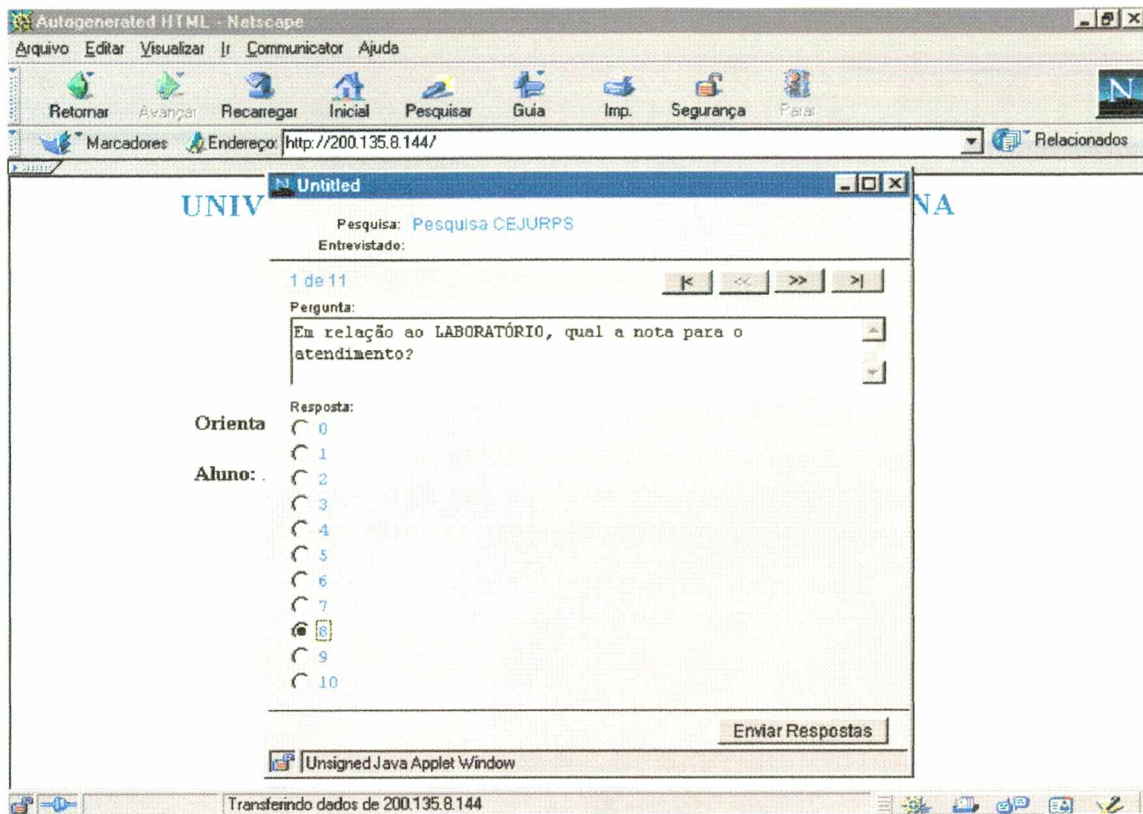


Figura 20. Respondendo a pesquisa.

CONCLUSÕES E RECOMENDAÇÕES

A pesquisa tradicional apropria-se da variável tempo em termos contextuais. As pesquisas estão territorializadas em momentos fixos de sua própria realização e seus resultados apresentam validade limitada pelo fato da variável tempo estar contemplada segundo a abordagem sincrônica, sendo que o pesquisador de campo é figura materializadora da percepção tradicional de espaço e tempo, como mostra a figura xx.

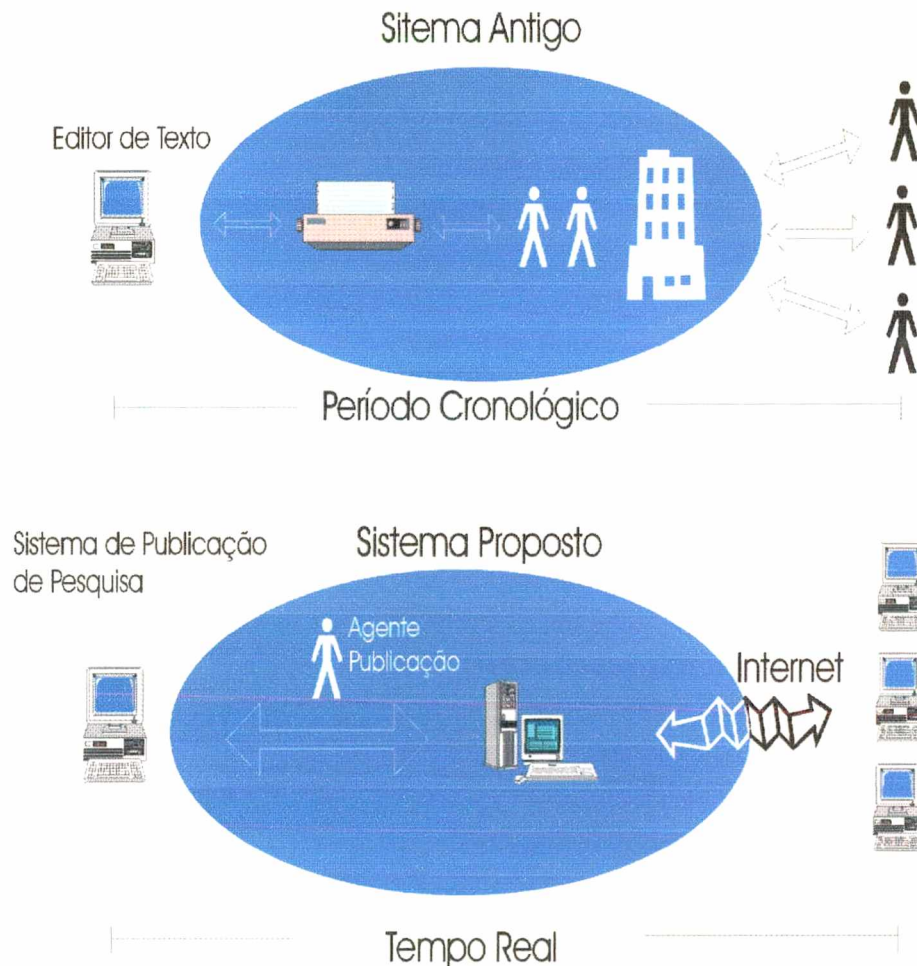


Figura 21. Sistema Antigo versus Sistema Proposto.

A pesquisa do modelo proposto apresenta outras configurações de procedimento e de análise de resultados. Aqui a variável tempo é apreendida em termos

processuais diante da possibilidade ativa de acumulação e análise de dados em tempo real de modo diacrônico. O tempo aqui é absorvido como variável processual e a validade de dados são complementadas em tempo real. A percepção diacrônica interfere diretamente na postura de análise de dados e elaboração de conclusões. Em virtude de teste-retestes, formas paralelas, interindicadores e no que diz respeito à validade.

No modelo proposto, as formas de acesso aos procedimentos investigativos não estão materializadas pela figura do pesquisador de campo, mas por formas de ingresso ao processo de pesquisa que reconstrói as percepções referenciais de espaço e tempo, corrompendo a idéia tradicional de espaço e tempo, simultaneamente.

São modelos distintos, modelos de ação de investigação diferenciados pela abordagem da variável tempo.

Portanto, resultados alcançados em períodos anteriores segundo o modelo proposto favorecem a análise diacrônica dos resultados mais atuais sem vácuos de períodos cronológicos de tempo.

O trabalho em questão, dificilmente terá seu fim alcançado, pois o abordado é uma pequena contribuição para o que há de se fazer.

Como trabalhos futuros, com objetivo de ampliar o mesmo, propões-se:

Trocar a linguagem Java para outra a qual deva-se observar a velocidade nas transmissões dos dados;

Fazer algoritmos de exportação de dados coletados para outras ferramentas estatísticas;

Criar a possibilidade de inserir gráficos tanto nas perguntas quanto nas respostas;

Estudar linguagem natural para que não haja duplicidade de perguntas;

Fazer um estudo dos dados coletados através das perguntas abertas e mistas, utilizando linguagem natural.

Desta forma, outros trabalhos podem ser iniciados a partir deste, de forma que o sistema em questão seja efetivamente utilizado.

REFERÊNCIAS BIBLIOGRÁFICAS

- Aamodt, A. e Plaza, E., Case-Based Reasoning: Foundational Issues, Methodological Variations, and Systems Approaches. *Artificial Intelligence Communications*, Vol. 7, No. 1, 1994.
- Aquilino, W.S. (1994). Interview Mode Effects in Surveys of Drug and Alcohol Use: A Field Experiment. *Public Opinion Quarterly*, 58, 210-240.
- Aquilino, W.S. and LoSciuto, L. (1990). Effects of Interview Mode on Self Reported DrugUse. *Public Opinion Quarterly*, 54, 362-395.
- Auer, K. (1995). Agents
<http://www.tip.net.au/~kauer/project/main.html> [janeiro, 1997].
- Augras, Monique. *Opinião Pública; Teoria e Pesquisa*. Petrópolis: Vozes, 1974
- Bachmann, D., Elfrink, J., & Vazzana, G. (1996). Tracking the progress of e-mail vs. snail-mail. *Marketing Research*, 8(2), 30-35.
- Belgrave, M. (1995). The Unified Agent Architecture: A White Paper [on-line].
http://www.ee.mcgill.ca/~belmarc/uua_paper.html [março, 1997].
- Batagelj, Z. and Vehovar, V. (1998). Technical and Methodological Issues in WWW Surveys. Paper presented at the annual meeting of the American Association for Public Opinion Research, St. Louis, May 1998.
- Bond, A. e GASSER, L. *An Analysis of Problems and Research in DAI. Readings in Distributed Artificial Intelligence*, Morgan Kaufmann, 1988.
- Bradburn, N. (1983). Response Effects. In *Handbook of Survey Research*, P. Rossi, J. Wright, and A. Anderson (eds.), 289-328. San Diego, CA: Academic Press.
- Bradshaw, J. *An Introduction to Software Agents*. In *Software Agents*, ed. J. M. Bradshaw, AAAI Press, 1997.
- Clayton, R.L. and Werking, G.S. (1998). Business Surveys of the Future: The World Wide Web as a Data Collection Methodology. Chapter 27 in Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls, W.L. II, and O'Reilly, J. (eds.), *Computer Assisted Survey Information Collection*. New York: Wiley.
- Colosi, Laura *The Layman's Guide to Social Research Methods*
<http://trochim.human.cornell.edu/tutorial/colosi/lcolosi1.htm> [janeiro, 2000]
- Comley, P. (1997). The Use of the Internet as a Data Collection Tool. Paper

- presented at the ESOMAR Annual Conference, Edinburgh, September 1997, <<http://www.sga.co.uk/esomar.html>>.
- Couper, Mick P. et al. A Comparison of Mail and E-mail for a Survey of Employees in U.S. Statistical Agencies. *Journal of Official Statistics*, Vol. 15, No. 1, 1999, pp. 39-56.
- Couper, M.P. and Nicholls II, W.L. (1988). The History and Development of Computer Assisted Survey Information Collection. In Couper, M.P., Baker, R.P., Bethlehem, J.,
- Clark, C.Z.F., Martin, J., Nicholls II, W.L., and O'Reilly, J. (eds.), *Computer Assisted Survey Information Collection*. New York: Wiley, 1-21.
- Dillman, D.A. (1978). *Mail and Telephone Surveys: The Total Design Method*. New York: Wiley.
- Dillman, D.A. (1991). The Design and Administration of Mail Surveys. *Annual Review of Sociology*, 17, 225±249.
- Durgan, Sally R. You're the Evaluator
<http://trochim.human.cornell.edu/tutorial/durgan/wic1.htm> [janeiro, 2000]
- Etizioni, O., Weld, D. S., Intelligent Agents on the: Fact, Fiction and Forecast. *IEEE Expert* 10(4): 44-49, 1995.
- Farley, S. R., *Mobile Agent System Architecture: A flexible alaternative to moving data and code to complete a given task*. Java Report. SIGS Publication, Inc. New York, NY, USA, 1997.
- Ferreira, Aurélio Buarque de Holanda, *Dicionário Aurélio da Língua Portuguesa*, Rio de Janeiro: Editora Nova Fronteira, 1994.
- Finin, T., WEBER, J. et al. DRAFT: Specification of KQML Agent Communication Language. The DARPA Knowledge Sharing Initiative External Interfaces Working Group,
<http://www.cs.umbc.edu/kqml/kqml.ps> [June 1993].
- Finin, T. (1997). *Agent Programming and Scripting Languages*.
<http://www.cs.umbc.edu/agents/technology/asl.shtml> [maio,1997]
- Foner, L. What is an Agent, Anyway? A Sociological Case Study. *Agents Memo*, Media Lab, MIT, pp.93-101, 1993.
- Franklin S. e Graesser, A. Is it an Agent, or just a Program?: A Taxonomy for Autonomous Agents. *Proceedings of the Third International Workshop on Agent Theories*, Springer-Verlag, 1996.
- Genesereth, M. R., Ketchpel, S. P. Software Agents. *Communications of the ACM*, July 1994, vol. 37 , n° 7, pg. 48-53, 147.

- Gentner, D. Structure Mapping – A Theoretical Framework for Analogy. *Cognitive Science*, Vol. 7, pp. 155-170, 1983.
- Gfroerer, J.C. and Hughes, A.L. (1991). The Feasibility of Collecting Drug Abuse Data by Telephone. *Public Health Reports*, 106, 384-393.
- Gilbert, D., Janca, P., IBM Intelligent Agents. IBM Corporation, Research Triangle Park, USA, 1996.
- Goguen, J. A. Fuzzy Sets in *Journal of Mathematical Analysis and application*, Vol 18, pp. 145-174, 1967.
- Hermans, B., Intelligent Software Agents on the Internet: na inventory of currently offered functionalith in the information society & a prediction of (near) future developments. Tilburg University, Tilburg, The Netherlands, july, 1996.
- Hewitt, C. Viewing Control Structures as Patterns of Passing Messages. *Artificial Intelligence* 8, Vol. 3, 1977.
- IBM, Open Blueprint: Intelligent Agent Resource Maneger. IBM Corporation, Document Number G325-6592-00, 1995.
- Kaufmann, A. Introduction a la Theorie des Sous-ensembles Flous. Masson : Paris, 1975.
- Kay, A. Computer Software. *Scientific American* 251 (3), pp. 53-59, September, 1984.
- Kearney, P. (1996). Personal Agents: A walk on the client side. [on-line] URL <http://www.sharp.co.uk/pk/unicom/Unicom.htm> [abril,1997]
- Kern, E., Uma Estrutura de agentes para o Processo de Licitação. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis SC, 1998.
- Kickert, Walter J. M. Fuzzy Theories on Decision-Making. Eindhoven University of Technology : Eindhoven ,1978.
- Kiesler, S. and Sproull, L.S. (1986). Response Effects in the Electronic Survey. *Public Opinion Quarterly*, 50, 402±413.
- Kittleson, M. J. (1995). An assessment of the response rate via the postal service and e-mail. *Health Values*, 18(2), 27-29.
- Klir, George J. and Yuan, Bo. Fuzzy Sets and Fuzzy Logic. Prentice Hall Prt : New Jersey, 1995.
- Kolodner, J. Reconstructive Memory, a Computer Model. *Cognitive Science*, Vol. 7, pp. 281-328, 1983.
- Kolodner, J. Case-Based Reasoning. Morgan Kaufmann, 1993.

- Kolodner, J. e Leake, D. A Tutorial Introduction to Case-Based Reasoning. Case-Based Reasoning: Experiences, Lessons, & Future Directions. Ed. David B. Leake, AAAI Press, 1996.
- Leake, D. B. Case-Based Reasoning: Experiences, Lessons, & Future Directions. AAAI Press, 1996.
- Leenw, E.D. and van der Zouwen, J. (1988). Data Quality in Telephone and Face-to-Face Surveys: A Comparative Meta-Analysis. In Telephone Survey Methodology, R. Groves, P. Biemer, L. Lyberg, J. Massey, W. Nicholls, and J. Waksberg, (eds.), 283-299. New York: Wiley and Sons
- Lenmon, B., Pynadath, D., Taylor, G. e Wray, B. Cognitive Architectures, <http://krusty.eecs.umich.edu/cogarch4/toc.html> , [1996].
- Mehta, R., & Sivadas, E. (1995). Comparing response rates and response content in mail versus electronic mail surveys. Journal of the Market Research Society, 37(4), 429-439.
- Mugo, Fridah W. Sampling In Research
<http://trochim.human.cornell.edu/tutorial/mugo/tutorial.htm> [janeiro, 2000]
- Mowbray, T. Essentials of Object-Oriented Architecture. Object Magazine, September 1995, pp. 28-32.
- Nissen, M. Intelligent Agents: A Technology and Business Application Analysis. <http://haas.berkeley.edu/~heilmann/agents/#exnum> , [November, 1995].
- Nwana, H., Software Agents: An Overview. Knowledge Engineering Review, Vol. 11, N. 3, pp. 1-40, September 1996.
- Ochard, D. (1996). Intelligent Agents [on-line].
<http://www.pacificspirit.com/Courses/Agents/27.htm> [novembro, 1996]
- Parker, L. (1992, July). Collecting data the e-mail way. Training and Development, 52-54.
- Ramos, M., Sedivi, B.M., and Sweet, E.M. (1998). Computerized Self-Administered Questionnaires (CSAQ). In Couper, M.P., Baker, R.P., Bethlehem, J., Clark, C.Z.F., Martin, J., Nicholls II, W.L., and O'Reilly, J. (eds.). Computer Assisted Survey Infor-mation Collection. New York: Wiley, 389±408.
- Riesbeck, C. e Schank, R. Inside Case-based Reasoning. Lawrence Erlbaum Associates, Institute for the Learning Sciences, 1989.
- Ross, Timothy J. Fuzzy Logic with Engineering Application. McGraw Hill : New York, 1995.
- Rymarchyk Gretchen K. Validity
<http://trochim.human.cornell.edu/tutorial/rymarchk/rymar2.htm> [janeiro, 2000]

- Shackle, G. L. S. *Decision, Order and Time in Human Affairs*. Cambridge University Press, New York, 1961.
- Schank, R. *Dynamic Memory: A Theory of Reminding and Learning in Computers and People*. Cambridge University Press, 1982.
- Schank, R. e Abelson, R. *Scripts, Plans, Goals and Understanding: An Inquiry Into Human Knowledge Structures*. Lawrence Erlbaum Associates, Institute for the Learning Sciences, 1977.
- Schuldt, B., & Totten, J. (1994). Electronic mail vs. mail survey response rates. *Marketing Research*, 6, 36-39.
- Schaefer, D. and Dillman, D. (1998). Development of a Standard E-mail Methodology: Results of an Experiment. *Public Opinion Quarterly*, 62, 378-397.
- Schuldt, B.A. and Totten, J.W. (1994). Electronic Mail vs. Mail Survey Response Rates. *Marketing Research*, 6 (1), 36-44.
- Smith, C.B. (1997). Casting the Net: Surveying an Internet Population. *Journal of Computer-Mediated Communication*, 3 (1), <<http://www.ascusc.org/jcmc/vol3/issue1/smith.html>>.
- Sosdian, C.P. and Sharp, L.M. (1980). Nonresponse in Mail Surveys: Access Failure or Respondent Resistance. *Public Opinion Quarterly*, 44, 396±402.
- Souza, E. M. S., *Uma Estrutura de Agentes para Assessoria na Internet*. Dissertação de Mestrado, Universidade Federal de Santa Catarina, Florianópolis SC, 1996.
- Spector, L. Automatic Generation of Intelligent Agent Programs. *IEEE Expert Intelligent Systems*, February 1997, pp. 3-4.
- Stone, P., Veloso, M., *Multiagent Systems: A survey from a Machine Learning Perspective*. Computer Science Department, Carnegie Mellon University, Pittsburg, PA, 1997.
- Sun, R., Commonsense reasoning with rules, cases, and connectionist models: A paradigmatic comparison. *Fuzzy Sets and Systems*, 82, 1996, pp. 187-200.
- Sycara K. e Pannu A. et al. Distributed Intelligent Agents. *IEEE Expert Intelligent Systems*, December 1996, pg. 36-46.
- Swoboda, W.J., MuÈhlberger, N., Weitkunat, R., and SchneeweiÙ, S. (1997). Internet Surveys by Direct Mailing. *Social Science Computer Review*, 15 (3), 242±255.
- Thiry, M., Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 1999.
- Trochim, William M., *Research Methods Knowledge Base*. Cornell University, 2000 <http://trochim.human.cornell.edu/kb/> [janeiro, 2000]

- Tse, A.C.B., Tse, K.C., Yin, C.H., Ting, C.B., Yi, K.W., Yee, K.P., and Hong, W.C.. (1995). Comparing Two Methods of Sending out Questionnaires: E-mail versus Mail. *Journal of the Market Research Society*, 37 (4), 441±446.
- Tse, A.C.B. (1996). Comparing the Response Rate, Response Speed and Response Quality of Two Methods of Sending out Questionnaires: E-mail vs Mail. Paper presented at the International Conference on Social Science Methodology, Essex, July 1996.
- Van Velthoven, G. Fuzzy Models in Personnel Management, paper presented et third Intenational Congress of Cybernetics and Systems, Bucharest, 1975.
- Watson, I. Applying Case-Based Reasoning: Techniques for Enterprise Systems. Morgan Kaufmann, 1997.
- Webber, R. L. Pesquisa Jurisprudencial Inteligente. Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Produção, Universidade Federal de Santa Catarina, 1998.
- Wooldridge, M. e Jennings, N. R. Intelligent Agents: Theory and Practice. *The Knowledge Engineering Review*, vol. 10, nº 2, pg. 115-152, 1995.
- Wooldridge, M. e Jennings, N. R. Software Agents. *IEEE Review* January 1996, pp. 17-20.
- Zadeh, L. A. Fuzzy Sets in Information and Control, Vol 8

BIBLIOGRAFIA

American Statistical Association

<http://www.amstat.org/publications/index.html> [março, 2000]

Ander-egg, Ezequiel. Introducción a las Técnicas de Investigación Social; Para Trabajadores Sociales. Buenos Aires: Humanitas, 1978.

Boyd Junior, Harper W., Westifall, Ralph. Pesquisas Mercadológicas: Textos e Casos. Rio de Janeiro: Getúlio Vargas, 1978.

Carmines, E.G. & Zeller, R.A. (1979). Reliability and validity assessment. Beverly Hills: Sage.

Christine B. Smith Naval Postgraduate School Monterey, CA Casting the Net: Surveying an Internet Population

<http://www.ascusc.org/jcmc/vol3/issue1/smith.html>[maio, 2000]

Cooper, C.L. & Mangham, I.L. (1971). T-groups: A Survey of Research. London: Wiley-Interscience.

Comley, Pete. The Use of the Internet as a Data Collection Method

<http://www.sga.co.uk/esomar.html> [março, 2000]

Cronbach, L.J. & Meehl, P.C. (1955). Construct validity in psychological tests. Psychological Bulletin, 52: 281-302.

FIND/SVP. (1995). The American Internet users survey.

<http://etrg.findsvp.com/index.html> [fevereiro, 2000]

Foner, L. (1996) What is an Agent, Anyway? A Social Case Study

<http://foner.www.media.mit.edu/people/foner/Julia/Julia-intro.html>

[maio, 1996]

Georgia Tech's Graphics, Visualization, and Usability Center (1996). GVU's WWW User Surveys [On-line]. Available URL: http://www.cc.gatech.edu/gvu/user_surveys/

Gilbert, D. Manny, A. (1996). Intelligent Agent Strategic [on-line]. URL <http://activist.gpl.ibm.com:81/WhitePaper/ptc.html> [novembro, 1996].

Groves, R.M. (1990). Theories and Methods of Telephone Surveys. Annual Review of Sociology, 16, 221-240.

- Hewson, C. M., Laurent, D., & Vogel, C. M. (1996). Proper methodologies for psychological and sociological studies conducted via the internet. *Behavior Research Methods, Instruments, & Computers*, 28, 186-191.
- Hinkin, T.R. & Schriesheim, C.A. (1989). Development and application of new scales to measure the French and Raven (1959) bases of social power. *Journal of Applied Psychology*, 74(4): 561-567.
- Hoelter, J.W. (1983). The analysis of covariance structures: Goodness-of-fit indices. *Sociological Methods and Research*, 11: 325-344.
- Hogan, R. (1969). Development of an empathy scale. *Journal of Consulting and Clinical Psychology*, 33, 307-316.
- Johnson, How Stable and Consistent Is Your Instrument?
<http://trochim.human.cornell.edu/tutorial/johnson/melody.htm> [janeiro, 2000]
- Journal of Official Statistics
<http://www.jos.nu/Articles/> [maio, 2000]
- Kaplan, Abraham. *A conduta na Pesquisa: Metodologia para as Ciências do Comportamento*. São Paulo: E. P. U. /EDUSP, 1975.
- Kehoe, C. M., & Pitkow, J. E. (1996). Surveying the territory: GVU's five WWW user surveys. *The World Wide Web Journal*, 1(3).
http://www.cc.gatech.edu/gvu/user_surveys/papers/w3j.html
- Kieley, J. M. (1996). CGI scripts: Gateways to world-wide web power. *Behavior Research Methods, Instruments, & Computers*, 28, 165-169.
- Kiesler, S., & Sproull, L. (1986). Response effects in the electronic survey. *Public Opinion Quarterly*, 39, 82-101.
- Lemay, L., Murphy, B. K., & Smith, E. T. (1996). *Creating commercial web pages*. Indianapolis, IN: Sams.net Publishing.
- Lissitz, R.W. & Green, S.B. (1975). Effect of the number of scale points on reliability: A Monte Carlo approach. *Journal of Applied Psychology*, 60: 10-13.
- Mann, Peter H. *Métodos de investigação Sociológica*. Rio de Janeiro: Zahar, 1970
- Marconi, Marina de Andrade e Pacatos, Eva M. *Técnicas de Pesquisa*. São Paulo: Atlas, 1990.
- Nielsen Media Research/CommerceNet. (1995). *The CommerceNet Nielsen Internet Demographics Survey*.
<http://www.nielsenmedia.com/> [fevereiro, 2000]
- O'Reilly Research. (1995). *Defining the Internet Opportunity*.
<http://www.ora.com/www/research/users/index.html> [dezembro, 1999]

- O'Reilly & Associates (1995). Defining the internet opportunity [On-line]. Available URL: <http://www.ora.com/research/users/index.html>
- Palmquist, J., & Stueve, A. (1996). Stay plugged in to new opportunities. *Marketing Research*, 8 (1), 13-15.
- Pardinas, Felipe. *Métodología y Técnica de Investigación en Ciencias Sociales: Introducción Elemental*. México: Siglo Veintiuno, 1977.
- Patrick, A. S., Black, A., & Whalen, T. E. (1995). Rich, young, male, dissatisfied computer geeks? Demographics and satisfaction from the National Capital FreeNet. In D. Godfrey & M. Levy (Eds.), *Proceedings of Telecommunities 95: The International Community Networking Conference* (pp. 83-107). Victoria, British Columbia, Canada: Telecommunities Canada.
<http://debra.dgbt.doc.ca:80/services-research/survey/demographics/vic.html>
[abril, 2000]
- Pitkow, J. E., & Recker, M. M. (1994). Using the web as a survey tool: Results from the second WWW user survey. *Journal of Computer Networks and ISDN Systems*, 27 (6).
http://www.cc.gatech.edu/gvu/user_surveys/papers/survey_2_paper.html
[março, 2000]
- Quarterman, J. S. (1996). Summary: Third Internet demographic survey (MIDS3).
<http://www3.mids.org/mn/603/ids3sum.html> [novembro, 1999]
- Richardson, Roberto J. e Colobodares. *Pesquisa Social: Métodos e Técnicas*. São Paulo: Atlas, 1985.
- Rosenfeld, P., Booth-Kewley, S., & Edwards, J. E. (1993). Computer-administered surveys in organizational settings: alternatives, advantages, and applications. *American Behavioral Scientist*, 36(4), 485-511.
- Rossi, P. H., Wright, J. D., & Anderson, A. B. (1983). *Handbook of Survey Research*. New York: Academic Press.
- Schmidt, W. C (1997). *Behavior Research Methods, Instruments & Computers*, 29(2), 274-279.
- Schmidt, W. C. (1996). World wide web survey research: Benefits, potential problems, and solutions. *Behavior Research Methods, Instruments, & Computers*, in press.
- Schmidt, W. C., Hoffman, R., & MacDonald, J. (1996). *Behavior Research Methods, Instruments, & Computers*, in press.
- Selltiz, C., et al *Métodos de Pesquisa nas Relações Sociais*. São Paulo: Herder, 1965

- Sproull, L. S. (1986). Using electronic mail for data collection in organizational research. *Academy of Management Journal*, 29, 159-169.
- Trochim, W.M. (1985). Pattern matching, validity, and conceptualization in program evaluation. *Evaluation Review*, 9(5), 575-604.
- Zimmermann, Hans Jürgen. *Fuzzy Set Theory – and its Applications*. Academic Publishers : Massachusetts, 1991.
- Young, Cheri A. Validity Issues in Measuring Psychological Constructs: The Case of Emotional Intelligence
<http://trochim.human.cornell.edu/tutorial/young/eiweb2.htm> [abril, 2000]
- Watson, M. & Greer, S. (1983). Development of a questionnaire measure of emotional control. *Journal of Psychosomatic Research*, 27(4), 299-305.
- William J. & Hatt, Paul K. *Métodos em Pesquisa Social*. São Paulo: Nacional, 1969.
- Witt, Aracy. *Metodologia de Pesquisa: Questionário e Formulário*. São Paulo: Resenha Tributária, 1975.

APÊNDICE I

TEORIA DE AGENTES

1.1 Introdução

Segundo Etzioni e Weld (1995) as tecnologias da informática têm aprimorado a habilidade de gerar, distribuir e armazenar informações e as ferramentas para localizá-las, filtra-las não tem acompanhado, sendo que Agentes Inteligentes são uma solução.

Agente Inteligente tem sido utilizado para determinar entidades de softwares que executam um conjunto de operações em nome do usuário ou de outro programa. Embora o termo Agente Inteligente inclua a palavra Inteligente, estas entidades não, necessariamente, precisem apresentar o comportamento inteligente, mas sim um determinado grau de autonomia e para tal empregam algum conhecimento dos objetivos ou desejos do usuário, IBM (1995).

Spector (1997) confirma a definição acima e acredita que uma característica positiva dos agentes é sua facilidade de implementação já que são sistemas autônomos e limitados, sendo possível gerar programas agentes automaticamente, programação evolutiva-genética, onde agentes seriam resultados de uma constante evolução.

Assim, Pattie Maes (1998) nos ensinam que agentes serão cada vez mais frequentes nas redes, pois estes oferecem diversos serviços como introduzir inteligência às interfaces, realizar comércio eletrônico e entre outras atuar como procuradores de usuários, tudo isto devido a sua versatilidade.

1.2 Histórico

O estudo da tecnologia de agentes teve origem em meados da década de 50, quando John McCarthy e Oliver G. Selfridge, ambos no MIT (Massachusetts Institute of Technology), tinham em mente um sistema que, dado um objetivo, poderia lidar com os detalhes das operações necessárias para cumpri-lo, pedir e receber conselhos em termos lingüísticos, caso houvesse problema. Um agente deveria ser um *Soft Robot*

vivendo e fazendo as suas ações dentro de um mundo de computadores segundo Kay (1984) apud Bradshaw (1997).

Hewitt (1977), na década de 70, definiu o conceito de um objeto auto-suficiente, interativo e com execução concorrente, o qual chamou de ator.

Com a disseminação de redes de computadores, segundo Bond (1988), emerge a área de Inteligência Artificial Distribuída (IAD), esta sendo um subcampo da Inteligência Artificial (IA). A IAD pode ser subdividida em duas áreas Thiry (1999):

- Sistemas Multi-Agentes (MAS – *Multi-Agent Systems*);
- Resolução Distribuída de Problemas (DPS – *Distributed Problem Solving*).

Os DPS, segundo Stone e Veloso (1997), trabalham com o gerenciamento de informações, os quais consistem em:

- Decomposição de tarefas: uma tarefa complexa é dividida em várias subtarefas e enviadas para processadores diferentes;
- Síntese de solução: os resultados de diferentes subtarefas são combinados.

Já os MAS, segundo Stone e Veloso (1997), usam princípios para construção de sistemas complexos, envolvendo múltiplos agentes e mecanismos para coordenação do comportamento inteligente de um conjunto de agentes, para obter a solução de um problema.

Já na década de 90, com a popularização da Internet, os agentes tornaram-se mais visados na comunidade científica e acadêmica. Estima-se que a principal fatia de mercado neste ano, ano 2000, será o mercado das telecomunicações, o que indica um potencial enorme em sistemas de mensagens, recuperação de informação, soluções customizadas, interfaces voltadas para o usuário e novos paradigmas de desktop e laptop segundo Chorafas (1998).

1.3 Conceitos de Agentes

Segundo Ferreira (1994) um agente é uma pessoa que trata de negócios por conta alheia. Voltado ao mundo computacional, um agente passa a ser um componente de software que mantém a função de tratar de negócios alheios.

Análogo ao que acontece na vida real, um agente de software precisa interagir com outros agentes durante as fases de sua jornada, seja para alcançar objetivos ou resolver problemas. Assim, este precisa ter um certo grau de autonomia a fim de tomar pequenas decisões, isto é, de livre-arbítrio.

Para Chorafas (1998), agentes são pequenos módulos autônomos de software que interpretam vários papéis na execução de uma meta ou pedido do usuário.

Segundo Bigus (1998) torna-se relevante enfatizar a capacidade de percepção e reação ao mundo externo que um agente deve apresentar. Agentes devem ser capazes de perceber o mundo físico ou virtual ao seu redor utilizando sensores. Uma parte fundamental da percepção é a habilidade de reconhecer e filtrar eventos esperados e tratar eventos não esperados. Agentes inteligentes agem enviando mensagens a outros agentes, ou chamando rotinas de API (*Application Program Interface*) ou serviços do sistema operacional diretamente.

Maes considera a importância da interface e comunicação dos agentes em seu trabalho e apresenta duas definições para caracterizar um agente inteligente. Na primeira definição, agente autônomo é um sistema que tenta alcançar uma série de metas em um ambiente complexo e dinâmico, podendo sentir e agir sobre ele [MAESa]. No segundo conceito Agentes são programas de computador que aplicam técnicas de inteligência artificial para auxiliar usuários nas tarefas baseadas por computador [MAESb].

Wooldridge e Jennings (1995) enfatizam os conceitos de autonomia, reatividade, pró-atividade e comunicação com o mundo exterior. Um agente é um sistema de computador autônomo, autocontido, reativo e pró-ativo, tipicamente com *locus* central de controle, que é capaz de comunicar com outros agentes através de alguma linguagem de comunicação entre agentes.

Além das definições citadas, obviamente, existem outras, mas é relevante observar que essas definições dão atributos comuns aos agentes como inteligência, autonomia, comunicabilidade, confiabilidade, cooperatividade, discurso, flexibilidade, confiabilidade, aprendizagem, mobilidade, persistência, personalização, planejamento, pró-atividade, reatividade, representabilidade e responsabilidade. Cada autor apresenta a sua idéia do que seriam os atributos ideais para um agente, de acordo com as características e as necessidades funcionais do mesmo.

1.4 Atributos dos Agentes

Os atributos de um agente caracterizam-no quanto a sua funcionalidade e o conjunto destes atributos é utilizado como uma forma para agrupar os agentes em classes ou tipologia. Na Tabela 1, apresentada por Kern (1998), pode-se observar os atributos mais relevantes.

Propriedade	Descrição	Observação	Referências
Aprendizagem	Acumulação de conhecimento pelo agente, que adapta o seu comportamento à novas situações.	Aprendizado pode ser: <ul style="list-style-type: none"> ▪ Unidirecional; ▪ Interativo. Também referenciada como Adaptatividade.	<ul style="list-style-type: none"> ▪ Wooldrid e Jennings (1996) ▪ Belgrave (1995) ▪ Auer (1995) ▪ Lenmom apud Souza (1996)
Autonomia	O agente pode ser apto a realizar ações preemptivas e independentes, que eventualmente poderão beneficiar o usuário.	Também referenciada como Independência.	<ul style="list-style-type: none"> ▪ Auer (1995) ▪ Wooldrid e Jennings (1996) ▪ Foner (1996)
Comunicabilidade	Capacidade do agente comunicar-se com outros agentes, pessoas, objetos e seu ambiente, a fim de completar suas tarefas e auxiliar os demais.	Também referenciada como Habilidade Social.	<ul style="list-style-type: none"> ▪ Wooldrid e Jennings (1996) ▪ Franklin e Graesser (1996) ▪ Belgrave (1995)

			Auer (1995)
Confiabilidade	Agentes devem servir as necessidades do usuário de maneira Segura, com veracidade e benevolência, não apresentando informações falsas e objetivos conflitantes.		<ul style="list-style-type: none"> ▪ Wooldridge e Jennings (1996) ▪ Auer (1995) ▪ Sycara et al (1996)
Cooperatividade	Capacidade dos agentes trabalharem em conjunto para concluírem tarefas mutuamente benéficas e complexas.	Também referenciada como Colaboração.	<ul style="list-style-type: none"> ▪ Sycara et al (1996) ▪ Gilbert apud Souza (1996)
Degradação Gradual	Capacidade do agente executar parte de uma tarefa quando existe incompatibilidade na comunicação ou domínio.		<ul style="list-style-type: none"> ▪ Foner (1996)
Discurso	Troca de informações entre usuário e agente, propiciando um melhor conhecimento de suas habilidades e intenções.		<ul style="list-style-type: none"> ▪ Foner (1996)
Flexibilidade	Habilidade do agente em escolher dinamicamente as ações e a seqüência de execução das mesmas.		<ul style="list-style-type: none"> ▪ Jennings (1995) ▪ Auer (1995) ▪ Franklin e Graesser (1996)
Inteligência	Conjunto de recursos, atributos e características as quais habilitam um agente a decidir quais ações tomar.	O raciocínio desenvolve-se em três cenários: <ul style="list-style-type: none"> ▪ Regras; ▪ Conhecimento ▪ Evolução artificial. 	<ul style="list-style-type: none"> ▪ Auer (1995) ▪ Gilbert e Janca (1996) ▪ Nissen (1995)
Mobilidade	Habilidade do agente em mover-se de uma máquina para outra, entre diferentes plataformas e sistemas, preservando seu estado interno.	Apresenta três graus: <ul style="list-style-type: none"> ▪ Estático; ▪ Roteiros móveis ▪ Móveis com estado 	<ul style="list-style-type: none"> ▪ Auer (1995) ▪ Belgrave (1995) ▪ Orchard (1996)

Persistência	Capacidade de manter um estado interno conciso através do tempo, o qual não é modificado ao acaso.	Também referenciado como Continuidade Temporal.	<ul style="list-style-type: none"> ▪ Belgrave (1995) ▪ Etzioni e Weld (1995)
Personalização	O usuário determina como o agente deve interagir, este aprende as características, preferências e hábitos de seus usuários.		<ul style="list-style-type: none"> ▪ Gilbert e Janca (1996) ▪ Kearney (1996)
Planejamento	Habilidade de sintetizar e escolher entre diferentes seqüências de ações pretendidas para alcançar seus objetivos.		<ul style="list-style-type: none"> ▪ Belgrave (1995) ▪ Auer (1995)
Pró-Atividade	Agentes são capazes de exibir comportamento oportunístico e orientado a objetos, tomando a iniciativa quando achar apropriado.		<ul style="list-style-type: none"> ▪ Wooldridge e Jennings (1995) ▪ Belgrave (1995)
Reatividade	Capacidade de perceber e responder as modificações que ocorrem no seu ambiente.	Também referenciado como “Sentir e Agir”.	<ul style="list-style-type: none"> ▪ Hermans (1996) ▪ Franklin e Graesser (1996)
Representabilidade	Representação do usuário em termos de suas ações e não mera existência.	Pode ser de dois modos: <ul style="list-style-type: none"> ▪ Passiva; ▪ Ativa. 	<ul style="list-style-type: none"> ▪ Wooldridge e Jennings (1995) ▪ Auer (1995)
Responsabilidade	O agente requisitado a realizar uma tarefa, é responsável pela sua conclusão.		<ul style="list-style-type: none"> ▪ Auer (1995)

Tabela 1. Atributos mais relevantes. Kern (1998).

1.5 Taxionomia dos Agentes

Franklin e Graesser (1996) e Nwana (1996) apresentam diferentes propostas de taxionomia de agentes, já que não existe um consenso entre os pesquisadores da área. Franklin e Graesser consideram os agentes autônomos como classe raiz para a hierarquia e Nwana coloca os agentes lado a lado na mesma hierarquia, cada um especializado de acordo com as características que possui. A Figura 1 apresenta a taxionomia proposta por Franklin e Graesser (1996), enquanto a Figura 2 apresenta a taxionomia de Nwana (1996).

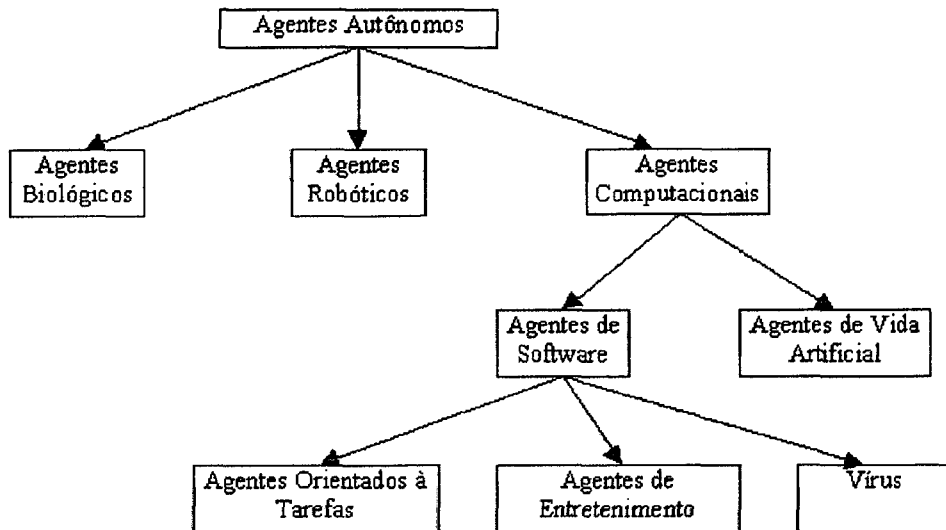


Figura 1. Taxionomia de Agentes. Franklin (1996).



Figura 2. Taxionomia de Agentes. Nwana (1996).

Kern (1998), Tabela 2, descreve a taxonomia de agentes segundo Nwana.

Agente	Descrição	Referência
Colaborativos	Enfatizam a autonomia e cooperação a fim de realizar tarefas a seus usuários. Suas características gerais incluem, além: habilidade social, responsabilidade e proatividade. Conseqüentemente, estes são capazes de agir racionalmente e autonomamente em ambientes multi-agentes abertos.	▪ Nwana (1996)
Interface	Enfatizam a autonomia e aprendizado a fim deste realizar as tarefas para seus usuários. Interagem com o usuário, recebendo especificações e fornecendo resultados.	▪ Nwana (1996) ▪ Sycara et al (1996)
Móvel	Processos capazes de navegar por uma rede, tal como a WWW, interagindo com host, reunindo informações no interesse de seu usuário e retornando, tendo executado os deveres ordenados.	▪ Nwana (1996) ▪ Farley (1997)
Informação	Executam o papel de gerenciador, manipulador e colecionar informações de diversas fontes distribuídas. Não existe um padrão para suas operações, podendo ter características variadas: estáticos	▪ Nwana (1996) ▪ Wooldridge e Jennings (1995)

	ou móveis, não-cooperativos ou sociais, e aprender ou não.	
Reativo	Agentes que não possuem modelos simbólicos internos do ambiente e atuam de acordo com um comportamento de estímulo/resposta.	▪ Nwana (1996)
Híbrido	Agentes que são constituídos pela combinação de duas ou mais filosofias de agentes. Adotados para tentar maximizar a potencialidade e minimizar as deficiências das várias filosofias.	▪ Nwana (1996)
Heterogêneos	Conjunto de dois ou mais agentes, os quais pertencem a duas ou mais classes de agentes diferentes. Um sistema de agente heterogeneo pode também conter um ou mais agentes híbridos.	▪ Nwana (1996)
Inteligente	Aspiração dos pesquisadores ao invés da realidade. Definições para estes foram descritas no item 1.3.	▪ Maes

Tabela 2. Taxonomia de agentes. Kern (1998).

1.6 Arquiteturas de Agentes

Uma arquitetura de agentes estabelece normas de comunicação entre agentes, exerce o agendamento de eventos, armazena a base de conhecimento, aciona a máquina de inferência e implementa o protocolo de comunicação.

Esta deve oferecer um ambiente para que se possa desenvolver e executar agentes em um meio seguro e robusto através de uma rede de computadores.

As arquiteturas requerem a execução de programas de compatibilização, tais como *Brokers* (Intermediadores) e *Runtime Environments* (Ambientes de tempo de execução). Esta característica obriga o usuário que deseja usufruir desta tecnologia a instalar estes programas em seus computadores. Enquanto não houver uma arquitetura padrão embutida em uma plataforma de software e hardware, este fato é uma realidade.

Conforme Mowbray (1995) apud Thiry (1999), procura-se estabelecer alguns conceitos que podem ser úteis para o desenvolvimento de uma arquitetura promissora:

- **Simplicidade:** idealizar a arquitetura e seus componentes de forma que seja fácil de entender, implementar e manter;
- **Funcionalidade:** selecionar uma arquitetura e ferramentas de desenvolvimento que focalizem aspectos específicos do problema a ser abordado;

- Expansividade: a arquitetura deve poder ser ampliada, uma vez que nem todas as necessidades futuras podem ser previstas em um primeiro momento;
- Isolamento ou portabilidade: uma arquitetura, para poder ser expansiva, deve possuir uma implementação portátil, evitando-se soluções não padronizadas.

1.6.1 Linguagem de Comunicação de Agentes

De acordo com Finin et al (1993) apud Thiry (1999) estabelecem que existem diversos níveis nos quais sistemas baseados em agentes devem interagir:

- Transporte: como agentes enviam e recebem mensagens;
- Linguagem: qual o sentido de mensagens individuais;
- Política: como os agentes estruturam conversações;
- Arquitetura: como conectar sistemas em concordância com protocolos existentes.

Assim, existem duas abordagens populares para o projeto de uma linguagem de comunicação descritas por Genesereth e Ketchpel (1994).

- Abordagem Procedural: a comunicação pode ser melhor modelada como troca de diretivas procedurais. Exemplos destas linguagens tem-se: TCL, Apple Events e Telescript;
- Abordagem Declarativa: a comunicação pode ser melhor modelada como a troca de sentenças declarativas, isto é, definições, suposições e outros.

APÊNDICE II

RACIOCÍNIO BASEADO EM CASOS

1.1 Introdução

O objetivo da Inteligência Artificial é representar o comportamento humano através de modelos computacionais. Os Sistemas Especialistas (SE) mostram-se um campo de pesquisa em IA de grande aplicabilidade devido às características de inteligência que estes sistemas apresentam.

No entanto, Watson (1997) aponta algumas dificuldades encontradas pelos pesquisadores da área de SE, tais como: o processo de recuperação do conhecimento é complicado, tornando sua implantação difícil; Tendem-se a se tornar lentos e incapazes de manipular grandes volumes de informação; e a inabilidade de aprender automaticamente torna o processo de manutenção difícil.

Assim, outras técnicas em IA surgiram de forma a preencher a lacuna deixada pelos SE's, estas baseadas na cognição humana. Para Sun (1996) apud Thiry (1999), o raciocínio é um aspecto fundamental da cognição humana. Este está envolvido em todos os tipos de processos cognitivos, variando do entendimento da linguagem ao processo de tomada de decisão. Neste sentido, Raciocínio Baseado em Casos (RBC) é uma técnica de IA que se utiliza a filosofia de aprender através de situações que acontecem no passado, através do aprendizado por experiências.

1.2 Histórico

Segundo Watson (1997), a história do Raciocínio Baseado em Casos, começa com a investigação do filósofo Wittgentein em 1953, onde observou que conceitos naturais, como mesas e cadeiras, são na verdade polimórficos e não podem ser classificados como um simples conjunto de facilidades suficientes e necessárias, mas

através de um conjunto casos (instâncias) que possuem similaridade, o que levou ao trabalho de Roger Schank em memória dinâmica, em 1982.

Roger Schank em seus trabalhos sobre memória dinâmica, manipulação de casos passados e padrões de situações, percebeu que poderiam ser aplicados à resolução de problemas e ao aprendizado, Schanck e Abelson (1997). Para Aamodt e Plaza (1994), este trabalho é uma das principais origens do RBC.

Baseado nos estudos de Roger Schank, Janet Kolodner, uma das precursoras do RBC, desenvolveu o Cyrus, o primeiro sistema a utilizar esta abordagem Kolodner, (1983). Este sistema possuía um modelo de memória baseado na organização de memória (*memory organization packets*), onde casos específicos com prioridades similares eram organizados em estruturas mais generalizadas. Este era um sistema baseado em perguntas e respostas sobre viagens e reuniões do secretário de estado americano Cyrus Vance. O Cyrus foi o deslanchar de muitos outros sistemas, Aamodt e Plaza (1994).

Segundo Watson (1997), atualmente, trabalhos envolvendo RBC tanto nos Estados Unidos quanto na Europa (Alemanha) têm se expandido, e o número de publicações vem crescendo, atingindo inclusive a área comercial.

1.3 Conceitos de RBC

Raciocínio Baseado em Casos é conceituado entre os autores de forma unanime, isto é, eles não divergem de que é uma técnica baseada em utilizar experiências passadas para resolver problemas análogos. A Tabela 1 apresenta alguns dos principais conceitos.

Fonte	Conceito
Kolodner (1983)	Modelo de raciocínio que envolve o entendimento de situações, a solução de problemas e o aprendizado, integrando esses aspectos com processos de memória.
Gentner (1983)	É um novo ramo de resoluções de problemas por analogia que utiliza a experiência para resolver novos problemas. Através do raciocínio por analogia, identifica determinados aspectos em novos e antigos problemas e procura utilizar as soluções encontradas para inferir uma nova solução.
Riesbeck e Schanck	RBC resolve novos problemas adaptando soluções que solucionaram problemas do

(1989)	passado.
Aamodt e Plaza (1994)	Sistemas capazes de utilizar um conhecimento específico de experiências passadas (casos) para resolver novos problemas.
Leake (1996)	RBC está baseado em lembranças, e que a memória de casos que caracterizam episódios específicos do passado.
Watson (1997)	Paradigma de resolução de problemas que envolve a aproximação entre o problema atual e um problema resolvido com sucesso no passado.
Webber (1998)	É uma técnica de Inteligência Artificial que produz aspectos da cognição humana para resolver problemas especialistas.
Thiry (1999)	É um método de raciocínio baseado na proposta de utilizar experiências passadas encapsuladas em estruturas de dados como a base para lidar com novas situações similares.

Tabela 1 – Definições de RBC

Os conceitos apresentados são de certa forma genérica demais, pois nem um deles por si só representa a magnitude desta técnica como um todo, pois há outras abordagens embutidas nesta técnica como: Raciocínio baseado em regras; Raciocínio baseado em exemplos; Raciocínio baseado em instâncias; Raciocínio baseado em memória; raciocínio baseado em analogias.

1.4 Fundamentação de RBC

Segundo Aamodt e Plaza (1994), a utilização de Raciocínio Baseado em Casos envolve quatro passos que devem ser observados, estes podem ser vistos na Figura 1.

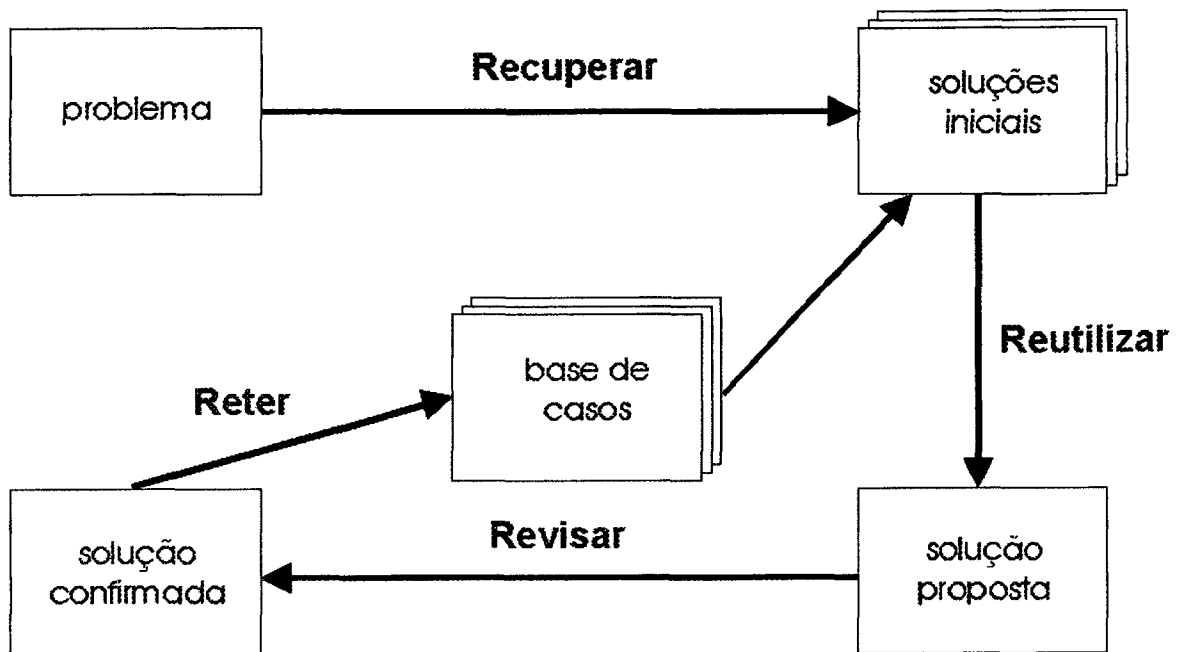


Figura 1 Ciclo RBC. Aamodt e Plaza (1994).

- Recuperar ou buscar na base casos um ou n casos comparando com a situação dada, isto é, o caso de entrada com os casos existentes na base de casos.
- Reutilizar ou aproveitar os casos similares existentes na base de casos para resolver o problema, isto é, o caso de entrada.
- Revisar ou avaliar a solução encontrada.
- Reter ou aprender a partir de um caso de entrada, isto é, armazenar o novo caso na base de casos.

1.4.1 Representação de Casos

De acordo com Webber (1998), a representação dos casos num sistema de RBC é essencialmente a representação do conhecimento. Há outros momentos em que algum conhecimento especialista é representado no sistema de RBC, entretanto, é nos casos que está representado o conhecimento que servirá para sugerir uma solução para o problema de entrada no sistema, a base de conhecimento está nos casos.

O problema de representação em RBC refere-se, fundamentalmente, em o que guardar de um caso, encontrando uma estrutura apropriada para descrever o conteúdo do caso e decidindo como a memória de casos deve ser organizada e indexada para uma efetiva recuperação e reutilização.

Segundo Watson (1997), há uma falta de consenso dentro da comunidade do RBC para definir exatamente que tipo de informação deveria estar em um caso. Entretanto, duas medidas podem ser tomadas para decidir o que deveria ser representado em casos: a funcionalidade e a facilidade da informação representada no caso.

Assim, dois fatores devem ser levados em consideração, a modelagem dos casos e a modelagem de memória. Para a modelagem de casos, segundo Webber (1998), estes são normalmente representados através de uma lista de atributos devidamente valorados. As características de um caso referem-se ao par atributo-valor. Na representação dos casos, dois são os componentes básicos: a descrição do problema; a descrição da solução.

A descrição do problema descreve o problema através da atribuição de características que descrevem o caso de entrada. As características descritivas podem ter a forma de nomes, números, funções ou textos, e servem para representar características, objetivos, metas, restrições, condições. Servem ainda para identificar o caso e são estas características que determinam a similaridade com outro caso.

Kolodner e Leake (1996), no que se refere a descrição das características descritivas, sugere duas diretrizes para decidir a relevância de incluir determinada descrição, onde a primeira sugere a inclusão de todos os aspectos que tenham contribuído para atingir a meta que é representada pelo caso e a segunda trata da inclusão das características que são normalmente utilizadas para descrever casos do tipo em estudo.

As descrições da solução são as características que descrevem a solução do caso e devem, basicamente, apontar a solução do problema de entrada e informar qual o resultado da aplicação desta solução ao mesmo.

Pode-se descrever a solução através de uma metodologia, esta deve mencionar a forma com que foi implementada e deve ser acompanhada de uma justificativa para a escolha desta opção.

Já para a modelagem de memória, é o conjunto de casos que englobam a base de conhecimento de um sistema de RBC, segundo Webber (1998).

Dentro do escopo de RBC, ao tratar-se de modelagem de memória, pode-se enfocar dois aspectos separadamente, o tipo de filosofia de representação que um dado sistema simula, que pode ser redes semânticas, memória episódica ou memória dinâmica ou o estilo de organização adotado para os casos, utilizado quando pretende-se um enfoque de implementação.

A filosofia de modelagem de memória refere-se ao tipo de memória utilizada para a representação do conhecimento. Um dos primeiros modelos de representação de conhecimento na IA foi o modelo de memória com uso de redes semânticas, o qual não é capaz de representar totalmente o conhecimento. O modelo posterior, a memória episódica, do psicólogo E. Tulving veio tentar complementar o modelo de redes semânticas.

Schank e Abelson (1977) desenvolveram o modelo de memória conceitual, dos *scripts* e posteriormente dos MOPs e o modelo de Memória Dinâmica.

Na memória dinâmica os casos são caracterizados pelos episódios aos quais estão associados e seus atributos são, não apenas os seus próprios, mas os atributos de suas abstrações e ligações que, juntos, modelam o contexto do caso.

Por outro lado, o estilo de organização dos casos na memória é escolhido com a natureza dos dados, com a forma de representação dos casos e com as outras etapas de desenvolvimento do sistema de RBC.

Os casos podem ser armazenados numa estrutura hierárquica ou numa estrutura plana de dados, ou ainda, em estruturas intermediárias que sejam originalmente planas mas que apontem para um nível de índices.

A indexação dos casos é um instrumento para orientar a similaridade, sendo um problema central do RBC, envolvendo a determinação dos tipos de índices que serão

utilizados pela etapa de recuperação como, por exemplo: índices preditivos ou explicativos; índices bastante abstratos para fazer um caso útil em uma variedade de situações futuras; índices bastante concretos a serem reconhecidos para futuros casos.

A indexação é uma forma de distinguir o que é importante sobre um caso e, os índices são a base para algoritmos de procura e as estruturas de organização utilizadas na base de casos, segundo Webber (1998).

1.4.2 Recuperação de Casos

A recuperação começa com uma descrição do problema, isto é, com um caso de entrada e termina quando um ou n caso é encontrado, sendo esta tarefa subdividida em: identificação das características:

- Informa ao sistema as características do caso atual;
- Casamento inicial ou similaridade recupera um conjunto de candidatos plausíveis;
- Busca é um processo mais elaborado de selecionar o melhor candidato entre os casos recuperados durante o casamento inicial;
- Seleção é o processo que gera conseqüências e expectativas de cada caso recuperado e tenta avaliar as conseqüências e justificar as expectativas.

Dentre estas tarefas, a similaridade é a essência do Raciocínio Baseado em Casos, uma vez que o fundamento do paradigma de RBC é solucionar um problema atual, caso de entrada, reutilizando a solução de uma experiência passada similar, segundo Webber (1998).

Para Aamodt e Plaza (1994), a avaliação da similaridade subdivide-se em similaridade sintática e semântica.

Similaridade sintática é a mais superficial, onde os atributos são comparados em termos de sua semelhança sintática. Pode-se citar algumas categorias como

sinônimos, ordinais, análise de perfil, clusterização e qualificadores, propostas por Witaker.

Já a similaridade semântica propõe-se uma avaliação mais profunda, tentando abranger o significado dos casos e comparando-os.

Métrica de Similaridade é uma função que mede analiticamente os graus de similaridade entre dois casos. Uma métrica é normalmente necessária em sistemas nos quais os casos são comparados um a um e a medida de sua similaridade é o meio de distinguir entre os casos candidatos similares e não similares. Assim, esta tem por objetivo dar um valor numérico à similaridade entre dois casos.

Existem dois aspectos fundamentais da recuperação, qual a similaridade entre os casos para cada dimensão e qual a importância de cada dimensão. Para modelar computacionalmente os diferentes graus de importância das características utiliza-se pesos ou algum mecanismo similar. Há ainda o conceito de *threshold* (limiar), que tem por objetivo limitar a quantidade de casos a serem recuperados. A Tabela 2 mostra alguns dos métodos de recuperação de casos.

Métodos	Descrição
Recuperação Indutiva (Inductive Retrieval)	Algoritmos de indução determinam quais características distinguem melhor o caso e gera uma estrutura tipo árvore de decisão para organizar os casos na memória.
Flat Memory, Serial Search	Os casos são armazenados sequencialmente em uma lista simples de array ou arquivo e a recuperação de casos se dão pela aplicação de uma função de casamento sequencialmente a cada caso do arquivo.
Shared Feature Networks (SFN)	Este método consiste do agrupamento de casos, passando a tratar apenas alguns pequenos conjuntos durante a recuperação.
Discrimination Networks (DN)	O algoritmo para achar itens na DN inicia no topo de uma árvore e a primeira questão é feita ao novo caso (caso de entrada), em seguida seleciona-se o nodo correspondente a sua resposta, continuando até que se encontre um item único no último nodo da árvore.
Redundant Discrimination Networks (RDN)	Este método tenta responder ao problema das informações perdidas. As pesquisas nas árvores são feitas em paralelo, sendo que em caso da não existência da resposta a uma questão em uma árvore, ocasiona a interrupção apenas da pesquisa desta árvore, continuando em outra árvore.
Flat Library, Parallel Search	Com uma máquina paralela "pesada", pode-se aplicar funções de casamento a todos os casos na ordem deles. Entretanto, quando trata-se o problema de

	indexação como um problema de prever sob quais circunstâncias um caso poderá ser recuperado de forma apropriada, o problema de indexação transforma-se em problema de casamento.
Hierarchical Memory, Parallel Search	Associam-se cada caso e cada nodo na hierarquia a um processador. Atribui-se pontos aos nodos, de forma que o nodo mais específico em cada parte da hierarquia tenha alguns pontos a mais que os nodos antecedentes.
Nearest Neighbour Retrieval	Consiste da identificação das características (atributos) essenciais para a solução do problema, sendo que estes atributos devem ser representados em algum sistema de coordenadas, de forma que se possa medir a distância entre o novo problema e os casos já existentes na memória de casos.

Tabela 2. Métodos de recuperação de casos.

1.4.3. Reutilização de Casos

A tarefa de reutilização de casos caracteriza-se por adaptar a solução armazenada de um caso recuperado às necessidades de um caso proposto.

Segundo Aamodt e Plaza (1994), a reutilização da solução dos casos recuperados no contexto dos novos casos tem foco em dois aspectos: as diferenças entre o caso passado e o caso corrente e que parte do caso recuperado pode ser transferida para o novo caso.

Em tarefas de classificação simples as diferenças são abstraídas de qualquer maneira (elas não são consideradas relevantes, enquanto as similaridades o são) e a solução do caso recuperado é transferida para o novo caso como solução para o mesmo. Entretanto, outros sistemas devem considerar as diferenças, já que a ligação entre os dois casos não é perfeita, e a reutilização faz-se através de um processo de adaptação que leva em consideração estas diferenças.

Assim, a tarefa de reutilização de casos está relacionada com o que se obteve, quando da aplicação a um caso real.

1.4.4. Revisão do Caso

Quando a solução de um caso gerada na fase de reutilização não está correta, nasce uma oportunidade para aprender a partir desta falha. Esta fase é chamada de revisão do caso, que consiste de duas tarefas: Avaliar e Reparar a solução.

Avaliar a solução do caso gerada pela reutilização e, se esta obteve sucesso, aprender (retenção do caso).

Reparar a solução do caso usando reconhecimento específico de domínio.

O conserto é um processo similar ao de adaptação, diferenciando-se pelo fato de que a adaptação começa com uma velha solução e o caso novo e os consertos são feitos com a solução deste caso novo.

Desta forma, torna-se o sistema hábil para fixar e examinar classes de falhas no futuro, podendo reconhecer situações onde existam falhas.

1.4.5. Retenção do Caso (Aprendizagem)

A aprendizagem é incorporar aquilo que é útil de reter do problema resolvido. Após a avaliação e possíveis reparos, efetua-se o aprendizado do sucesso ou das falhas da solução proposta.

Pode-se empregar a aprendizagem em sistemas de RBC em nível dos casos e da base de casos. Podendo as bases de casos ser estendidas através de processos incrementais de aprendizagem, uma vez que a tarefa e o projeto do sistema permitam isto.

De acordo com Leake (1996), a base de casos pode crescer com novos casos a partir de um pequeno conjunto de casos somente. A geração destes novos casos se dá através de novos casos informados pelos usuários ou a partir de uma fonte externa.

Na aprendizagem em nível dos casos, mantém-se no caso o registro de seu desempenho ao ser utilizado. Desta forma, tanto sucessos como fracassos são informados, incrementando o conhecimento e as lições embutidas no caso. Tal registro

também serve para prevenir o usuário com relação às possíveis conseqüências de seu uso. Assim, compensa-se a inclusão de informações no caso, evitando a reutilização de sugestões menos favoráveis, o que resulta no aumento da qualidade da recuperação.

APÊNDICE III

Tomada de Decisão Difusa

1.1 Discussão Geral

A tomada de decisão é, sem dúvida, uma das atividades mais fundamentais, segundo Klir (1995), para os seres humanos, pois no nosso dia a dia nos deparamos com uma variedade de ações e algumas vezes temos que decidir como agir.

O início dos estudos sobre a tomada de decisão pode ser datado do século XVIII, quando vários estudos foram feitos na França com relação à eleição e escolha social. Desde estes estudos iniciais, a tomada de decisão desenvolveu-se como um campo respeitável e rico de pesquisa. A literatura corrente sobre tomada de decisão é enorme e é em sua grande parte baseada em teorias e métodos desenvolvidos neste século.

O assunto tomado de decisão é como o nome sugere, o estudo de como decisões podem ser tomadas e como elas podem tornar-se melhores e mais bem sucedidas. Isto é, o campo se concentra, em geral, em ambas as teorias descritivas e normativas. Muito do foco no desenvolvimento do campo tem sido na área de gerenciamento, a qual o processo de tomada de decisão é de importância-chave para funções tais como controle de inventário, investimento, ações pessoais, desenvolvimento de novos produtos, e alocação de recursos, entre outras. A tomada de decisão por si só é amplamente definida para incluir qualquer escolha ou seleção de alternativas, e é, portanto, importantíssima em muitos campos como o das Ciências Sociais, Naturais e Engenharia.

Aplicações de conjunto difuso dentro do campo de tomada de decisão, em grande parte, consiste de “fuzzyficações” da teoria de tomada de decisão clássica. Enquanto a tomada de decisão sob condições de risco tem sido modelada pelas teorias de decisão probabilísticas e teorias de jogos, a teoria de decisão difusa atenta para o *vagueness* (algo que é vago) e a não, especificidade inerente na formulação humana de preferências, restrições, e objetivos.

Tomadas de decisão clássicas geralmente tratam com um conjunto de estados da natureza alternativos (resultados), um conjunto de ações alternativas que estão disponíveis para o tomador de decisões, uma relação indicando o estado ou resultado a ser esperado de cada ação alternativa, e, finalmente, uma função utilidade ou objetivo, que ordena o resultado de acordo com a “desejabilidade”.

Uma decisão é dita ser tomada sob condições de “certeza” quando o resultado para cada ação pode ser determinado e ordenado precisamente. Neste caso, a alternativa que leva ao resultado, produzindo a utilidade mais alta, é escolhido. Isto é, o problema de tomada de decisão torna-se um problema de otimização, o problema de maximização da função utilidade. Uma decisão é tomada sob condições de risco, por outro lado, quando o único conhecimento disponível concerne no resultado que consiste de sua distribuição de probabilidade condicional, uma para cada ação. Aqui, o problema de tomada de decisão torna-se um problema de otimização de maximização da utilidade esperada. Quando as probabilidades dos resultados são desconhecidas, ou ainda não relevantes, e os resultados para cada ação são apenas caracterizados aproximadamente, dizemos que as decisões são tomadas sob incerteza. Este é o principal domínio da tomada de decisão difusa.

Tomada de decisão sob incerteza é talvez a categoria mais importante dos problemas de tomada de decisão, tão bem caracterizado pelo economista britânico Shackle (1961):

Num mundo predestinado, uma decisão seria ilusória; num mundo de pré-conhecimentos perfeitos, uma decisão seria vazia; num mundo sem ordem natural, seria fraca. Nossa atitude intuitiva para viver, implica em decisões não ilusórias, não vazias, e nem fracas... Já que a decisão neste sentido exclui a previsão perfeita e a anarquia na natureza, deve ser definida como uma escolha em face a incerteza limitada.

Isto indica a importância da teoria dos conjuntos difuso na tomada de decisão.

Diversas classes de problemas de tomada de decisão são normalmente reconhecidas. De acordo com um critério, problemas de decisão são classificados como aqueles envolvendo um tomador de decisão única e aqueles envolvendo diversos tomadores de decisão.

Aquelas classes de problemas são referenciadas como tomada de decisão individual e estas tomadas de decisão multipessoal. De acordo com outro critério, distinguimos problemas de decisão como que envolvem uma otimização simples de uma função de utilidade, uma otimização sob restrições, ou uma otimização sob critérios de múltiplos objetivos. Além disso, a tomada de decisão pode ser feita em um estágio ou iterativamente em diversos estágios.

Não tentaremos cobrir a tomada de decisão difusa compreensivelmente sob todos os aspectos, ao invés disso, queremos levar o espírito de tomada de decisão difusa como aplicado a várias classes de problemas de decisão.

1.2 Tomada de Decisão Multicritério

Em problemas de decisão multicritérios, alternativas relevantes são avaliadas de acordo com um número de critérios. Cada critério inclui uma ordenação particular de alternativas, e precisamos de um procedimento pelo qual construímos uma ordenação global de preferências. Existe uma similaridade visível entre estes problemas de decisão e problemas de tomada de decisão multipessoal. Em ambos os casos, ordenações múltiplos de alternativas relevantes são envolvidos e tem que ser integradas em uma ordenação de preferências global. A diferença é que as ordenações múltiplas representam cada preferência de diferentes pessoas ou taxas baseadas em diferentes critérios.

Para efeito deste trabalho o número de critérios em uma tomada de decisão multicritérios é sempre assumido como finito. Assumimos também que o número de alternativas consideradas é também finito. Desta forma apresentaremos duas formas de agregação de critérios, plausíveis de solução para o problema exposto no Capítulo 4 desta tese. A primeira é apresentada por Kauffmann (1975) e a segunda descrita por Klir (1995).

1.2.1 Aplicação de Kauffmann

Segundo Kickert (1978) um tipo de problema de tomada de decisão que é diferente, mas de certo modo relacionado à tomada de decisão por muitas pessoas

(multipessoal), é a tomada de decisão multicritérios. Esta situação é aquela na qual todas as alternativas no conjunto de escolhas podem ser avaliadas de acordo com um certo número de critérios. O problema é construir um procedimento de avaliação para enfileirar o conjunto de alternativas em ordem de preferência. O problema de decisão multicritérios é normalmente descrito como:

- Um conjunto de alternativas $A = \{a_1, \dots, a_n\}$
- Um conjunto de n critérios, denotados por $1, 2, \dots, n$

n conjuntos em ordem de preferência O_1, \dots, O_n nos quais O_k apresenta-se como a ordenação das alternativas de acordo com o k -ésimo critério:

$$O_k: A \times A \rightarrow \{0,1\}$$

Resultando no conjunto ordenado $\{a_{k1}, a_{k2}, \dots, a_{kn}\}$ que é uma permutação de $A = \{a_1, \dots, a_n\}$.

- Um conjunto de pesos $W = \{W_1, \dots, W_n\}$ onde W_k denota a importância do critério k na avaliação das alternativas.

O problema de decisão multicritérios a ser esboçado aqui é o apresentado por Kauffmann (1975) no seu terceiro livro de sua série sobre conjuntos difusos, capítulo 87. A abordagem começa com uma definição abstrata de um critério, fator de definição de toda sua abordagem:

Um critério $C(x)$ num conjunto X é definido como aquele que estabelece uma estrutura neste conjunto.

Segundo Kickert (1978), esta estrutura pode ser definida pelo mapeamento de $X \times X$ em $\{0,1\}$, que dá um relacionamento binário que pode representar uma ordenação, uma treliça (lattice), etc. Uma Segunda definição desta estrutura é representá-la pelo mapeamento de X num conj L que possua uma configuração de ordenação, lattice, etc. Esta definição posterior é precisamente a definição de um conjunto L-Fuzzy segundo (Goguen, 1967). Assim, pela definição, um conjunto L-Fuzzy, constitui um critério em X . Representando todos os possíveis conjuntos (e subconjuntos) L-Fuzzy em X por L^X , podemos afirmar que

$$C(x) \in L^X.$$

Se confinarmos o conjunto L ao intervalo fechado $[0,1]$, obteremos conjuntos fuzzy no sentido usual de Zadeh (1965).

A noção de pesos – diferentes níveis de importância dos vários critérios – pode também ser embutida na teoria dos conjuntos difusos. Consideramos um conjunto suporte X e tomamos $L = [0,1]$, então o conjunto de critérios de ordem M será um subconjunto de L^X :

$$M = \{C^1(x), C^2(x), \dots, C^m(x)\} \subset L^X$$

Com

$$C^i(x) \in L^X \quad i=1,2,\dots,m.$$

Dado M e um conjunto Π de m pesos $\Pi = \{p_1, p_2, \dots, p_m\}$, $p_i \in [0,1]$, então o conjunto fuzzy $\sim M$ em M definido por

$$\begin{aligned} \mu_{\sim M}(C^i(x)) &= p_i && \text{se } C^i \in M \\ &= 0 && \text{se } C^i \notin M \end{aligned}$$

Neste caso é chamado “pesagem de ordem do critério”

Se, além disso, $\sum_{i=1}^m p_i = 1$ a pesagem será chamada de “normal”

Note que:

$$\sim M \in L^{(L^X)},$$

O problema da agregação do critério é representado por Kauffmann (1975) da seguinte maneira:

Tomamos X como sendo um conjunto (finito) e m um conjunto de conjuntos L-Fuzzy em X : $M = \{A_1, A_2, \dots, A_m\}$ com $A: X \rightarrow L$ (e.g. $L = [0,1]$). Então o problema consiste em encontrar uma ordenação destes m conjuntos L-Fuzzy.

Em Kauffmann (1975) este problema de agregação é sub dividido em duas situações: cada L possui propriedades cardinais (escala desconhecida além de 0 e 1), ou

L apenas possui propriedades ordinais (escala ordenada). Apresentamos aqui o método mais próximo do clássico.

Assumimos X sendo $X = \{x_1, x_2, \dots, x_n\}$ e supomos que um deles tem m conjunto difuso $M = \{A_1, \dots, A_m\}$ em X , caracterizados por suas funções de pertinência $\mu_{A_i}(x_j)$, $j=1, \dots, n$; $i=1, \dots, m$. Assumimos um conjunto de pesos $\Pi = \{p_1, \dots, p_n\}$. Então calculamos:

$$v(A_i) = \sum_{j=1}^n \mu_{A_i}(x_j) p_j$$

Este $v(A_i)$ induzirá a uma ordenação nos conjunto difuso A_i .

Ao invés de considerar este método *per se*, nós primeiro daremos um exemplo de uma aplicação que usa uma versão mais simples desta abordagem de agregação.

1.2.1.1 Exemplo

Em Van Velthoven (1975), métodos de multicritérios difuso são aplicados a problemas de gerenciamento de pessoal. O problema específico considerado é o recrutamento e a seleção de pessoal. Esta seleção de candidatos é executada pela comparação dos perfis dos candidatos P_1, P_2, \dots, P_n com o perfil P_0 requerido tendo um certo número de critérios C_1, C_2, \dots, C_m conforme tabela abaixo:

	1.1.1.1 C_1	1.1.1.2 C_2	C_m
P_0	P_0^1	$P_0^2 \dots$	P_0^m
P_1	P_1^1	$P_1^2 \dots$	P_1^m
P_n	P_n^1	$P_n^2 \dots$	P_n^m

Na abordagem não difusa o intervalo do candidato j para o critério $i: p_j^i$, é 0 ou 1 e vários métodos de avaliação de *rankings* pode ser usado, por exemplo:

$$q^j = \sum_{i=1}^m p_i^0 p_i^j$$

Que expressa o nº de critérios requeridos que são satisfeitos.

2- Vetor Produto com Vetor Complementar

$$q^j = \sum_{i=1}^m p_i^0 p_i^j + \sum_{i=1}^m (1-p_i^0)(1-p_i^j)$$

Que expressa o nº de características em comum com o perfil requerido

3- Distancia Complementar de Hamming

$$q^j = 1 - \frac{1}{m} \sum_{i=1}^m |p_i^0 - p_i^j|$$

Esta é a versão normalizada da anterior no caso de perdis não fuzzy.

4- Índice de Similaridade

$$q^j = 1 - \frac{\sum_{i=1}^m |p_i^0 - p_i^j|}{m + \sum_{i=1}^m p_i^0 p_i^j}$$

Que expressa a similaridade do perfil requerido.

Existe uma extensão direta deste método de seleção de recrutamento num método difuso. Os perfis são considerados agora como conjuntos difuso P_i no mesmo conjunto suporte $\{C_1, \dots, C_m\}$.

$$P_i = \{ C_1 | \mu_1^i; C_2 | \mu_2^i \dots \dots C_m | \mu_m^i \} \quad \mu_j^i \in [0,1].$$

Os procedimentos de seleção podem ser facilmente estendidos para este caso difuso. Contudo, deve-se notar que os procedimentos “Vetor produto” e “Vetor Produto com Vetor Complementar” perdem sua significância porque não indicam números muito grandes (*no longer indicates numbers*). O único procedimento de seleção restante e valioso é a “Distância Complementar de Hamming”.

$$q^j = 1 - \frac{1}{m} \sum_{i=1}^m |\mu_i^0 - \mu_i^j|$$

E o Índice de Similaridade

$$q^j = 1 - \frac{\delta [P_0; P_j]}{m + [P_0 \cap P_j]} = 1 - \frac{\sum_{i=1}^m |\mu_i^0 - \mu_i^j|}{m + \sum_{i=1}^m \mu_i^0 \wedge \mu_i^j}$$

Um método adicional que pode ser usado para classificar os candidatos é o “*outranking*”: o perfil do candidato P_i está fora da fila P_j se:

$$c_{ij} \geq c \quad e$$

$$d_{ij} > d$$

Onde c_{ij} é o número relativo das funções de pertinência produzindo

$$|\mu_i - \mu_0| \leq |\mu_j - \mu_0|$$

e

$$d_{ij} = \sup [|\mu_j - \mu_0| - |\mu_i - \mu_0|] \quad \text{se } d > 0 \text{ e contrário } 0.$$

O pedaço anterior da expressão é o c requerido e o posterior é o nível d , o menor é o número de candidatos que não estão fora da fila (non-outranked).

De acordo com Van Velthoven (1975) a abordagem de conjunto difuso divide-se em 4 passos fundamentais:

- 1 Definir um conjunto suporte adequado (conjunto referencial) para as características fuzzy, por exemplo, o suporte de sete elementos:

Altura (alto, médio, pequeno, pequeno);

Sexo (M,F);

Inteligência (muito inteligente, inteligente).

- 2 Definir as funções de pertinência do perfil requerido P_0 . Por exemplo, $P_0 = \{\text{alto}/0.8; \text{médio}/0.7; \text{pequeno}/0.0; \text{M}/1.00; \text{F}/0.00; \text{muito inteligente}/0.1; \text{inteligente}/0.7\}$

- 3 Encontra as funções de pertinência dos perfis candidatos P_1, \dots, P_m . De acordo com Van Velthoven (1975), este modo pode ser feito pelos métodos psicométricos (*psychometric*) e pela *fuzzificação* das características resultantes.

- 4 Calcula os índices de seleção para obter a classificação final.

DISCUSSÃO

A razão pela qual a aplicação de Van Velthoven (1975) foi apresentada depois da apresentação do método multicritérios fuzzy de Kauffmann (1975) é que no

problema de recrutamento de pessoal não são usados pesos para o critério e os procedimentos de avaliação propostos não eram tão simples como o cálculo de:

$$v(A_i) = \sum_{j=1}^n \mu_{A_i}(x_j) P_j$$

Como proposto por Kauffmann (1975) existe uma similaridade importante entre ambos os métodos: o enfileiramento de uma alternativa A_i de acordo com um critério X_j é representado por um valor de pertinência difusa $\mu_{A_i}(x_j)$, $\in [0,1]$ (ou μ_j^i no caso dos perfis). Naquele sentido, o enfileiramento difuso parece com um coeficiente de enfileiramento convencional, no qual o um conjunto difuso é normalizado no intervalo $[0,1]$.

Esta característica de enfileiramento difuso é bastante surpreendente segundo Kickert (1978) quando vemos o que Van Velthoven (1975) propôs usando “alto”, “médio” e “pequeno”, etc. Como conjuntos suporte os quais determinam valores de pertinência, geralmente termos como estes são considerados exemplos padrão de conjuntos difuso. O que esperamos de coeficientes de enfileiramento difuso é que estes coeficientes sejam preferencialmente conjunto difuso a uma função de pertinência única. Este é certamente o caso onde o enfileiramento pode ser dado apenas em termos de “bom”, “muito bom”, etc. É difícil imaginar que tipo de fuzzyficação é significativo quando o tomador de decisão ainda está apto a designar um coeficiente de avaliação numérico.

1.2.2 Aplicação de Klir

Para a aplicação deste método tomamos $X = \{x_1, x_2, \dots, x_n\}$ e $C = \{c_1, c_2, \dots, c_m\}$ sendo um conjunto de alternativas e um conj de critérios caracterizando uma situação de decisão, respectivamente. Então, a informação básica envolvida em tomadas de decisão multicritério pode ser expressa pela matriz.

$$R = \begin{array}{c} \\ \\ \\ \\ \\ \\ \\ \\ C_m \end{array} \begin{array}{ccccc} X_1 & X_2 & \dots & X_n \\ r_{11} & r_{12} & \dots & r_{1n} \\ r_{21} & r_{22} & \dots & r_{2n} \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ r_{m1} & r_{m2} & \dots & r_{mn} \end{array}$$

Assumimos primeiro que todas as entradas da matriz sejam n^os reais em [0,1], e que cada entrada r_{ij} expressa o grau para qual o critério c_i é satisfeito pela alternativa x_j ($i \in N_m, j \in N_n$). Então R pode ser visto como uma representação da matriz de uma relação fuzzy em $C \times X$.

O que pode acontecer é que, ao invés da matriz R com entradas em [0,1], uma matriz alternativa $R' = [r'_{ij}]$, cujas entradas são números reais arbitrários, seja inicialmente dada. Neste caso, R' pode ser convertida na matriz R desejada pela fórmula.

$$r_{ij} = \frac{r'_{ij} - \min_{j \in N_n} r'_{ij}}{\max_{j \in N_n} r'_{ij} - \min_{j \in N_n} r'_{ij}}$$

Para todo $i \in N_m$ e $j \in N_n$.

A abordagem mais comum pra problemas de decisão multicritério é convertê-los em problemas de decisão de um critério único. Isto é feito encontrando um critério global, $r_j = h(r_{1j}, r_{2j}, \dots, r_{mj})$, que para cada $x_j \in X$ é um agregado de valores $r_{1j}, r_{2j}, \dots, r_{mj}$ para os quais os critérios individuais c_1, c_2, \dots, c_m são satisfeitos.

Um exemplo do problema de decisão multicritério é o problema de recrutamento e seleção de pessoal. Neste problema particular, a seleção de condições de um dado conjunto de indivíduos, ditos x_1, x_2, \dots, x_n , é guiado pela comparação de perfis candidatos com o perfil requerido em termos de dados critérios c_1, c_2, \dots, c_m . Isto resulta na matriz R . As entradas r_{ij} de R expressam, para cada $i \in N_m$ e $j \in N_n$, o grau para o qual o candidato x_j ajusta-se com o perfil requerido em termos do critério c_i .

Um operador freqüentemente empregado é a média pesada (*weighted average*)

$$r_j = \frac{\sum_{i=1}^m \omega_i r_{ij}}{m + \sum_{i=1}^m \omega_i} \quad (j \in N_n)$$

Onde w_1, w_2, \dots, w_m são pesos que indicam a importância relativa dos critérios c_1, c_2, \dots, c_m . Uma classe de possíveis agregações pesadas é dada pela fórmula:

$$r_j = h(r_{1j}^{\omega_1}, r_{2j}^{\omega_2}, \dots, r_{mj}^{\omega_m})$$

Onde h é um operador de agregação e w_1, w_2, \dots, w_m são pesos.

Considere agora uma situação mais geral na qual as entradas da matriz R são números fuzzy \tilde{r}_{ij} em R^+ , e pesos são especificados em termos de números difusos \tilde{w}_i em $[0,1]$. Então, usando as operações de adição e multiplicação difuso, podemos calcular a média pesada \tilde{r}_j pela fórmula

$$\tilde{r}_j = \sum_{i=1}^m \tilde{\omega}_i \tilde{r}_{ij}$$

Já que números difusos são não linearmente ordenados, um método de enfileiramento é necessário para ordenar os n ºs fuzzy resultantes $r_1, \tilde{r}_2, \dots, \tilde{r}_n$.