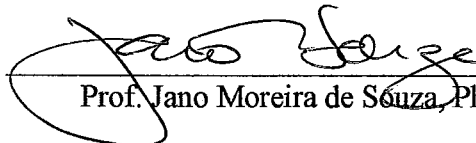


CEJ: UM AMBIENTE PARA A DEFINIÇÃO E EXECUÇÃO DE PROCESSOS
CUSTOMIZÁVEIS DE PUBLICAÇÃO CIENTÍFICA

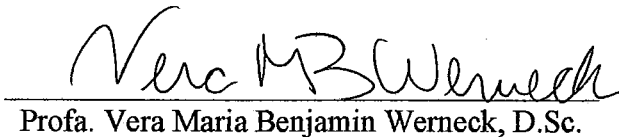
Daniel Serrão Schneider

TESE SUBMETIDA AO CORPO DOCENTE DA COORDENAÇÃO DOS
PROGRAMAS DE PÓS-GRADUAÇÃO DE ENGENHARIA DA UNIVERSIDADE
FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS
NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE MESTRE EM CIÊNCIAS EM
ENGENHARIA DE SISTEMAS E COMPUTAÇÃO.

Aprovada por:


Prof. Jano Moreira de Souza, Ph.D.


Prof. Sérgio Palma da Justa Medeiros, D.Sc.


Profa. Vera Maria Benjamin Werneck, D.Sc.

RIO DE JANEIRO, RJ – BRASIL

DEZEMBRO DE 2004

SCHNEIDER, DANIEL SERRÃO

CEJ - Um ambiente para a definição e
execução de processos customizáveis de
publicação científica [Rio de Janeiro] 2004

XIV, 153 p., 29,7 cm (COPPE/UFRJ, M.Sc.,
Engenharia de Sistemas e Computação. 2004)

Tese – Universidade Federal do Rio de
Janeiro. COPPE

1. Infra-estruturas do conhecimento científico
2. Sistemas de Anotações na Web.
3. Sistemas de *Workflows* customizáveis
4. Web Semântica

I. COPPE/UFRJ II. Título (série).

AOS MEUS AVÓS TERESA E JOSÉ

AGRADECIMENTOS

Primeiramente a Deus, por toda a força que me proporcionou durante a caminhada desta longa e árdua estrada do mestrado, e com quem compartilho todos os momentos de felicidade.

A duas pessoas maravilhosas que são os meus pais: Ligia Maria Schneider e Ricardo Luiz Schneider. Agradeço igualmente aos meus irmãos Adriana e André, aos meus avós Teresa e José, à minha tia Magaly e primo Pedro, e toda a minha família, e peço desculpas por ter me privado do convívio mais intenso com vocês em função deste trabalho de pesquisa.

A todos os colegas da minha turma do mestrado. O convívio de vocês foi muito importante ao longo desta jornada.

Aos meus amigos novos e antigos Gabriel Schneider, Camilla Peres, Cristiano Bretas, Joana Melo, Priscila Iague e à Branca Duboc. Um agradecimento especial e carinhoso à Cristina Soares, uma pessoa compreensiva que tive oportunidade de conhecer em 2004, e que me apoiou muito durante este último ano do mestrado.

À Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro durante os primeiros dois anos do mestrado.

Aos professores e comandantes José Roberto Blaschek e Francisco Pereira Neto, por todas as oportunidades que me foram oferecidas nos projetos da Marinha ao longo de quase seis anos; à comandante Célia pelo carinho com que me tratou durante todo esse tempo; ao comandante Giancarlo Cuquel pela confiança que depositou em mim nos últimos anos, e à qual eu tenho me empenhado em retribuir; um agradecimento especial ao comandante Glauco Monteiro de Castro, meu amigo e atualmente companheiro de equipe na Marinha, por todo o apoio que me foi dado para a conclusão deste trabalho de tese. As experiências trocadas foram muito gratificantes. Finalmente, um agradecimento aos colegas com quem tive oportunidade de trabalhar nos projetos da Marinha ao longo desses anos, sobretudo ao professor Geraldo Zimbrão.

A cinco pessoas que eu tive oportunidade de conhecer durante o período do mestrado, e que contribuíram indiretamente com este trabalho de pesquisa: Raquel Coelho Gomes, Frederico

Pontes, Emerson Morais, Carlos Ribeiro da Cunha e, especialmente, Carlos Augusto Garcia, o "Guto".

A todas as pessoas que torceram pelo sucesso deste trabalho. Em particular, gostaria de agradecer aos tios Lúcia e Hélder, e especialmente às minhas duas tias "Ana": Ana Maria Amaral e Anna Marya Schneider.

Um agradecimento também ao pesquisador alemão Alexander Hars, autor do livro "*From publishing to knowledge networks*", obra que forneceu muitos subsídios a este trabalho de tese.

A todas as pessoas que colaboraram nos experimentos realizados com o ambiente CEJ: ao professor Schneider, à Adriana Vivacqua, à Jonice Oliveira e a todos os alunos das duas turmas de graduação que forneceram numerosas contribuições para esta pesquisa.

Gostaria de agradecer imensamente aos professores co-orientadores Geraldo Bonorino Xexéo e Sérgio Palma da Justa Medeiros, pelas numerosas contribuições que foram dadas neste trabalho de pesquisa. Ao professor Xexéo por ter assentado os principais alicerces deste trabalho, e ao professor Palma por ter colocado um foco mais consistente nesta pesquisa, sobretudo por ter aberto um novo horizonte na minha vida - o da publicação científica - que resultou na minha viagem para a conferência da China.

Ao carinho da Patrícia Leal, sempre disposta a ajudar. Agradeço também a todos os professores que tiveram alguma participação na minha educação, começando pelo ensino fundamental, passando pelo ensino médio, pela graduação e finalmente o mestrado.

À professora Vera Werneck por ter aceitado participar da banca de avaliação desta monografia.

A todas as pessoas que contribuíram direta ou indiretamente com este trabalho.

Por último, um agradecimento mais do que especial ao professor e orientador que eu muito admiro, e que hoje ocupa o papel mais importante na minha vida acadêmica e profissional. Àquele que me abriu inúmeras portas desde o ano de 1999 quando o conheci: a porta dos projetos da COPPETEC, do curso de Mestrado, a porta para o meu trabalho de pesquisa, e até a porta para uma excursão à China! Ao professor que esteve sempre disposto a dialogar, a discutir idéias e a abrir novos caminhos: Jano Moreira de Souza.

Resumo da Tese apresentada à COPPE/UFRJ como parte dos requisitos necessários para a obtenção do grau de Mestre em Ciências (M.Sc.)

CEJ: UM AMBIENTE PARA A DEFINIÇÃO E EXECUÇÃO DE PROCESSOS
CUSTOMIZÁVEIS DE PUBLICAÇÃO CIENTÍFICA

Daniel Serrão Schneider

Dezembro/2004

Orientador: Jano Moreira de Souza

Programa: Engenharia de Sistemas e Computação

Um número crescente de portais, incluindo-se as revistas científicas eletrônicas, bibliotecas digitais, enciclopédias eletrônicas e laboratórios de colaboração, têm-se dedicado ao conhecimento científico nos últimos anos. Estes portais têm sido denominados *infra-estruturas de conhecimento científico* na literatura. Apesar do avanço aparente, em que as revistas científicas tradicionais estão migrando para a Web, as publicações se tornam digitais e algumas atividades do processo de publicação são automatizadas, uma análise mais atenta sobre essas infra-estruturas indica que o modelo de publicação científica permanece essencialmente inalterado. Nesse contexto nasceu a proposta do projeto denominado *Configurable Electronic Journal* (CEJ). Trata-se de um ambiente para a construção de revistas eletrônicas, que permite a definição e execução de processos de publicação configuráveis e extensíveis. Um dos principais objetivos desta proposta é de aplicar padrões e tecnologias abertas, sobretudo as da Web Semântica, no projeto de um ambiente inovador para a definição e execução de processos de publicação flexíveis. Um dos componentes que o ambiente CEJ provê consiste em um gerenciador de *workflows* dinâmicos e customizáveis, podendo acomodar diferentes estilos de publicação e adaptar-se a diferentes processos e disciplinas do conhecimento.

Abstract of Thesis presented to COPPE/UFRJ as a partial fulfillment of the requirements for the degree of Master of Science (M.Sc)

CEJ – AN ENVIRONMENT FOR THE DEFINITION AND EXECUTION OF
CUSTOMIZABLE SCIENTIFIC PUBLICATION PROCESSES

Daniel Serrão Schneider

December/2004

Advisor: Jano Moreira de Souza

Department: System and Computation Engineering

An increasing number of sites including scientific electronic journals, digital libraries, electronic encyclopedias and collaboratories have been dedicated to scientific knowledge in the recent years. These sites have been called scientific knowledge infrastructures in the literature. Despite the apparent progress, in which traditional scientific journals are migrating to Web, new forms of communication emerge, and some activities of the publication process are automated, a closer look at these sites indicates that the scientific publishing model remains essentially unchanged. In this context was born the CEJ project, which is a solution for generating configurable and extensible electronic journals. One of the main goals of this work is to apply open source technologies, including W3C Semantic Web technologies, in the design of an innovative environment for the definition of publication processes. CEJ's customizable workflow can accommodate different styles of publication, suitable to be used in different processes and disciplines.

SUMÁRIO

Agradecimentos.....	iv
Sumário.....	viii
Lista de Figuras.....	xii
Lista de Tabelas.....	xiv
Capítulo 1 – Introdução.....	15
1.1 - Motivação.....	15
1.2 - Descrição do problema e objetivos a atingir.....	18
1.3 - Contexto da tese.....	20
1.4 - Organização do trabalho de tese.....	21
Capítulo 2 – Infra-estruturas do conhecimento científico.....	22
2.1 - Considerações iniciais.....	22
2.2 - O processo tradicional de publicação.....	23
2.2.1 - Descrição do processo.....	24
2.2.2 - Descrição dos objetos envolvidos.....	25
2.2.3 - Descrição dos aspectos organizacionais.....	26
2.2.4 - A revisão pelos pares.....	27
2.2.5 - Problemas com o processo de publicação científica.....	28
2.2.5.1 - Lentidão do processo.....	28
2.2.5.2 - Crise dos seriais.....	29
2.2.5.3 - Revisão pelos pares fechada, binária e uma única vez.....	29
2.3 - Transição para as novas formas de publicação.....	31
2.3.1 - A mídia eletrônica.....	32
2.3.2 - O papel dos repositórios eletrônicos (eprints).....	32
2.3.3 - O impacto da tecnologia da informação.....	34
2.4 - Infra-estruturas de conhecimento científico.....	41
2.4.1 - Categorias de infra-estruturas de conhecimento.....	43
2.4.2 - Revistas eletrônicas - modelos emergentes de publicação científica.....	45

2.4.2.1 - Publicação sem revisão (non-refereeing).....	46
2.4.2.2 - O modelo da revista BMJ.....	46
2.4.2.3 - O modelo da revista BBS.....	47
2.4.2.4 - O modelo da revista Psycology.....	48
2.4.2.5 - O modelo da revista PubMed Central.....	49
2.4.2.6 - O modelo da revista JIME.....	50
2.4.2.7 - O modelo da Física de Alta Energia.....	52
2.4.2.8 - O modelo hipotético DJ.....	54
2.5 - Características das infra-estruturas de conhecimento científico baseadas na Web.....	55
2.5.1 - Modo de interação.....	55
2.5.1.1 - CSCW e Groupware.....	56
2.5.1.2 - Mecanismos de Awareness.....	58
2.5.1.3 - Sistemas de workflows.....	60
2.5.2 - Estruturação do conhecimento.....	62
2.5.3 - Apresentação.....	63
2.5.4 - Gerenciamento de uma infra-estrutura.....	65
2.5.5 - Tecnologia.....	67
2.6 - Sumário e Conclusões.....	69
Capítulo 3 – Sistemas de anotações baseados na Web.....	70
3.1 - Considerações iniciais.....	70
3.2 - Características dos sistemas de anotações.....	71
3.3 - Classificação dos sistemas quanto à arquitetura.....	72
3.4 - Infra-estrutura de apoio às anotações.....	73
3.5 - Análise comparativa de Sistemas de Anotações na Web.....	75
3.5.1 - CaMILE.....	75
3.5.2 - DocReview.....	77
3.5.3 - CritLink.....	78
3.5.4 - CoNote.....	79

3.5.5 - ComMentor.....	80
3.5.6 - Annotea.....	82
3.6 - Quadro comparativo das ferramentas analisadas.....	84
3.7 - Sumário e conclusões.....	85
Capítulo 4 – O ambiente CEJ.....	87
4.1 - Considerações iniciais.....	87
4.2 - Propostas do ambiente CEJ.....	87
4.2.1 - Modo de interação.....	87
4.2.2 - Estruturação do conhecimento.....	91
4.2.3 - Apresentação.....	92
4.2.4 - Gerenciamento de uma infra-estrutura.....	93
4.2.5 - Tecnologia.....	94
4.3 - Análise e Projeto do ambiente CEJ.....	96
4.3.1 - O paradigma orientado a objetos.....	96
4.3.2 - Modelagem conceitual.....	97
4.3.3 - Uso do padrão de projeto MVC “Model 2”.....	97
4.3.4 - Arquitetura do ambiente CEJ.....	100
4.4 - Descrição dos componentes da arquitetura do ambiente CEJ.....	101
4.4.1 - Componente de “Navegação e Anotação”.....	101
4.4.1.1 - Navegando sobre um documento.....	101
4.4.1.2 - Criando, visualizando e removendo anotações.....	103
4.4.1.3 - Filtrando anotações.....	105
4.4.2 - Componente de “Busca de documentos e anotações”.....	105
4.4.3 - Componente de “Gerência do ciclo de vida”.....	106
4.4.4 - Componente de “Gerenciamento de Workflows”.....	109
4.4.5 - Componente de “Notificações personalizadas”.....	111
4.4.6 - Componente de “Administração e Configuração”.....	112
4.5 - Implementação e protótipo.....	114
4.5.1 - Java e o framework Struts.....	114

4.5.2 - Implementação do “Gerenciador de Workflows”	117
4.5.3 - Implementação da API de geração de PDF.....	120
4.5.4 - Portal de desenvolvimento do ambiente CEJ.....	120
4.5.4.1 - Navegação sobre os artefatos de desenvolvimento e compilação automática das classes Java.....	121
4.5.4.2 - Acesso ao SGBD.....	122
4.5.5 - Performance do ambiente CEJ.....	123
Capítulo 5 – Estudos de caso com o ambiente CEJ.....	125
5.1 - Considerações iniciais.....	125
5.2 - Modelos propostos para a validação do ambiente CEJ.....	125
5.2.1 - Modelo 1: Uma revista de fluxo contínuo com publicação instantânea.....	125
5.2.2 - Modelo 2: Uma revista baseada no convite com fase de proposta.....	128
5.3 - Descrição do experimento I.....	129
5.4 - Descrição do experimento II.....	129
5.5 - Sumário e conclusões.....	133
Capítulo 6 – Conclusões e trabalhos futuros.....	134
6.1 - Contribuições desta dissertação.....	136
6.1.1 - Quatro mitos da publicação científica.....	137
6.2 - Trabalhos futuros.....	139
6.3 - Considerações finais.....	140
Referências bibliográficas.....	142
APÊNDICE A - Framework de características proposto por HARS (2003).....	148
APÊNDICE B - Modelagem conceitual do ambiente CEJ.....	149

LISTA DE FIGURAS

Figura 2.1 - O modelo tradicional de publicação científica.....	24
Figura 2.2 - Diagrama de classes de uma revista científica tradicional.....	25
Figura 2.3 - Categorias de infra-estruturas do conhecimento.....	42
Figura 2.4 - Ciclo de vida de um artigo submetido ao JIME.....	51
Figura 2.5 - Exemplo de documento sob revisão no JIME.....	52
Figura 2.6 - Classificação de Groupware.....	58
Figura 2.7 - Quatro tipos de processos.....	61
Figura 3.1 - Screenshot de uma instância do WebCaMILE.....	76
Figura 3.2 - Screenshot de uma instância do DocReview.....	77
Figura 3.3 - Screenshot de uma instância do CoNote.....	79
Figura 3.4 - Screenshot de uma instância do CoNote.....	80
Figura 3.5 - Screenshot de uma instância do ComMentor.....	81
Figura 3.6 - Screenshot de uma instância do Annotea.....	83
Figura 4.1 - Arquitetura MVC “Model 1”.....	99
Figura 4.2 - Arquitetura MVC “Model 2”.....	99
Figura 4.3 - Componentes da arquitetura do ambiente CEJ.....	100
Figura 4.4 - Modos de navegação sobre um documento.....	101
Figura 4.5 - Criando uma anotação localizada no ambiente CEJ.....	103
Figura 4.6 - Navegando sobre uma anotação localizada no ambiente CEJ.....	104
Figura 4.7 - Filtros de visualização das anotações no ambiente CEJ.....	105
Figura 4.8 - Conversor de HTML para XML do ambiente CEJ.....	107
Figura 4.9 - Definição de processos de publicação no ambiente CEJ.....	110
Figura 4.10 - Execução de processos de publicação no ambiente CEJ.....	111
Figura 4.11 - Configuração de templates de exibição e de tipos de anotação.....	113
Figura 4.12 - Configuração dos casos de uso do ambiente CEJ.....	114
Figura 4.13 - Arquitetura MVC “Model 2” de implementação do CEJ.....	116

Figura 4.14 - Esforço x flexibilidade na solução de um workflow.....	118
Figura 4.15 - Portal de desenvolvimento do ambiente CEJ.....	121
Figura 4.16 - Navegação sobre artefatos de desenvolvimento no CEJ.....	122
Figura 4.17 - Acesso ao SGBD pelo portal de desenvolvimento do CEJ.....	123
Figura 4.18 - Tela de “Estatísticas do ambiente CEJ”	124
Figura 5.1 - O modelo de fluxo contínuo com publicação instantânea.....	126
Figura 5.2 - Uma revista baseada no convite e com uma fase de proposta.....	128
Figura 5.3 - Escolha de um grupo de trabalho.....	130
Figura 5.4 - Escolha do tema de trabalho.....	131
Figura 5.5 - Documentos submetidos à revista ES2004.....	132
Figura B.1 - Modelo de classes geral do ambiente CEJ.....	150
Figura B.2 - Modelo conceitual da máquina de workflow do CEJ.....	152

LISTA DE TABELAS

Tabela 3.1 – Comparação das ferramentas de anotações para a Web.....	84
Tabela A.1 – Características das infra-estruturas do conhecimento científico.....	148

Capítulo 1

Introdução

1.1 Motivação

A publicação científica é o processo através do qual o conhecimento é descoberto, refinado, certificado, distribuído e preservado para os pesquisadores, professores, estudantes e o público em geral. A história do conhecimento científico está intimamente associada ao progresso da tecnologia da informação. Neste contexto, a invenção da escrita, por volta de 3500 A.C. na Mesopotâmia, é considerada a primeira revolução na tecnologia da informação. A escrita possibilitou o armazenamento bem como a transmissão do conhecimento, marcando o nascimento da primeira grande civilização. A próxima revolução na tecnologia da informação ocorreu por volta de 100 D.C. na China, quando Ts'ai Lun inventou o papel. A terceira revolução ocorreu em meados de 1400 D.C., quando Gutenberg inventou a prensa tipográfica, possibilitando o nascimento da imprensa e dos livros impressos. A tecnologia da impressão reduziu significativamente os custos e o tempo necessário para reproduzir manuscritos, o que era realizado até então manualmente.

No meio científico, vivemos nos dias de hoje um momento de transição da era da publicação impressa de Gutenberg, para a era da publicação eletrônica e da Internet – denominada por HARNAD (1991) de *Galáxia Pós-Gutenberg*. "a quarta revolução no que se refere à produção de conhecimento". Por publicação eletrônica entendemos qualquer mecanismo de distribuição de informação em uma forma que possa ser acessada e visualizada pelo computador e que utiliza recursos digitais para adquirir, armazenar e transmitir informação de um computador para outro (STANEK & PURCELL, 1995). Comparada às revoluções anteriores da tecnologia da informação, a revolução atual, baseada nas redes de computadores e na mídia digital, além de reduzir sensivelmente os custos de armazenamento e reprodução da informação, oferece funcionalidades inteiramente novas: "A informação pode ser transmitida instantaneamente para lugares distantes; o conhecimento armazenado eletronicamente pode ser modificado mesmo após o momento de sua publicação" (HARS, 2003).

O sistema de publicação constituído ao longo de diversas décadas, e baseado na invenção de *Gutenberg* - o papel impresso - se tornou extremamente caro para ser sustentado pela comunidade acadêmica, o que em parte se deve ao aumento exagerado do número, volume e dos custos das publicações. Por outro lado, o advento recente da *World Wide Web* em conjunto com a difusão da mídia eletrônica precipitaram uma revolução nas formas de comunicação científicas, motivando a redefinição do modelo de publicação científica à luz das novas tecnologias.

Um número crescente de portais, incluindo-se aí revistas eletrônicas, bibliotecas digitais, enciclopédias eletrônicas e laboratórios de colaboração, têm-se dedicado ao conhecimento científico nos últimos anos. Estes portais têm sido denominados “infra-estruturas de conhecimento científico” na literatura, ou ainda “redes de conhecimento científico”, termo preferido por alguns autores. Revistas e conferências no mundo inteiro passaram a adotar, nos últimos anos, a submissão e revisão de artigos de forma eletrônica. Em muitos casos, a versão final do artigo é publicada na Web como um documento hipertexto, ficando disponível ao público em geral. Popularizaram-se também os chamados repositórios abertos de literatura, espalhados por todo o mundo, que oferecem o texto completo dos artigos a um clique do mouse. Neste contexto, de acordo com ODLYZKO (2001), leitores e pesquisadores passam a ter à sua disposição um “rio de conhecimento”, que garante o acesso à informação através de uma diversidade de fontes.

Apesar deste avanço aparente, em que as revistas científicas tradicionais estão migrando para a Web, as publicações se tornam digitais e algumas atividades do processo de publicação são automatizadas, uma análise mais atenta sobre essas infra-estruturas indica que o modelo de publicação científica permanece essencialmente inalterado. Como regra geral, ao invés de se repensar o modelo tradicional de publicação, a mídia eletrônica e as tecnologias abertas da Internet têm sido utilizadas apenas para melhorar as práticas tradicionais, automatizando-se algumas atividades, e disponibilizando a literatura no formato eletrônico.

Na era da Internet e da globalização, um requisito fundamental para o avanço da pesquisa científica é, sem dúvida alguma, a propagação rápida da informação revisada pelos pares. A atividade de revisão pelos pares é ainda considerada um dos pilares do processo de publicação científica, responsável pelo controle de qualidade da literatura

publicada no meio científico. A captação de recursos para a pesquisa científica e o progresso de um pesquisador em sua carreira acadêmica ainda estão associados intimamente à publicação em um espaço pequeno de tempo.

A publicação eletrônica na Web possibilita a disseminação rápida e em larga escala da informação científica, sem limitação de espaço e a um custo baixo, indo de encontro aos requisitos apresentados. Além disso, a Web proporciona uma visão mais rica dos documentos, incluindo recursos multimídia (arquivos sonoros, filmes etc), apontadores para trechos do próprio documento ou para outros documentos, inclusive bi-direcionais, anotações sobre o texto e demais recursos do documento, linhas de discussão etc. Vale ressaltar ainda que a possibilidade de comunicação eletrônica entre pesquisadores, através de diversos mecanismos como o correio eletrônico, listas e fóruns de discussão, comunicação peer-to-peer, entre outros, tende a impulsionar a comunicação científica de forma globalizada e mais abrangente.

Segundo ODLYZKO (1999), a evolução natural do processo de publicação no meio científico para o formato eletrônico é motivada por dois fatores principais: a redução potencial dos custos inerentes ao processo de publicação, e a oferta de novas e atrativas funcionalidades, que, via de regra, não seriam possíveis com a mídia do papel. Todavia, em função da inércia no meio acadêmico e dos papéis envolvidos no atual sistema de publicação, esta mudança tende a se dar de forma lenta (ODLYSZO, 1999). Conforme mencionamos, as revistas científicas tradicionais, e mesmo as que surgiram após o advento da Internet, têm insistido na preservação dos principais mecanismos do modelo antigo. Ao invés de rediscutirem o modelo de publicação à luz das novas tecnologias, elas relutam em experimentar as novas formas e possibilidades proporcionadas pela mídia eletrônica, como forma de garantir o *status quo*. Por outro lado, existem evidências concretas de que o processo de comunicação científica vive um momento de rápida evolução (ODLYZKO, 2001), com o surgimento de formas inovadoras de comunicação baseadas nas tecnologias da Internet.

É bem verdade que algumas poucas revistas e portais vêm experimentando recentemente formas alternativas, em que o processo de publicação foi de certo modo repensado, e algumas inovações foram conseguidas neste sentido. É o caso da revista *BMJ*, na área da Medicina, da revista *Psicology* na área da Psicologia, e do revolucionário

modelo de publicação na área da *Física de Alta Energia*, entre outros. Estes exemplos serão analisados detalhadamente no capítulo 2 desta monografia, que trata das infra-estruturas emergentes de conhecimento científico. No entanto, se levarmos em conta um universo de 24000 revistas publicadas no mundo inteiro (HARNAD, 2004), estes casos constituem a exceção, e não a regra.

1.2 Descrição do problema e objetivos a atingir

Dentro do contexto descrito na seção anterior, em que o modelo tradicional de publicação científica encontra-se em derrocada, e a transição para as novas formas de publicação em andamento, nasceu a proposta do projeto denominado *Configurable Electronic Journal* (CEJ). Trata-se de um ambiente para a construção de revistas eletrônicas, que permite a definição e execução de processos de publicação configuráveis e extensíveis. Um dos principais objetivos desta proposta é de aplicar padrões e tecnologias abertas, sobretudo às da Web Semântica, incluindo-se aí a *eXtended Markup Language* (XML), *XPointer* e tecnologias de anotação, no projeto de um ambiente inovador para a definição e execução de processos de publicação flexíveis.

No capítulo 2 desta monografia, uma análise teórica das revistas tradicionais, seguida por uma análise experimental das revistas eletrônicas emergentes, dentro do contexto das chamadas “infra-estruturas de conhecimento”, terá como objetivo atingirmos algumas metas básicas: 1) a identificação de problemas existentes no processo tradicional de publicação que poderiam ser equacionados e resolvidos com a implementação de outros modelos; 2) a definição de um ou mais modelos de publicação, que poderiam ser alvo de experimentação no ambiente CEJ, com o objetivo de reduzir ou minimizar os problemas identificados; 3) o projeto de uma revista eletrônica, identificando-se que características seriam desejáveis em uma infra-estrutura de conhecimento, e de que forma a tecnologia da informação poderia concretizá-las.

Um dos componentes idealizados para o ambiente CEJ foi um gerenciador de *workflows* dinâmicos e customizáveis, podendo acomodar diferentes estilos de publicação e adaptar-se a diferentes processos e disciplinas do conhecimento. Neste sentido, seria possível ao corpo editorial de uma revista configurar as atividades envolvidas no processo

de publicação, os atores envolvidos em cada atividade, qual a forma de discurso (aberta, fechada, duplamente fechada) a ser empregada nas atividades de revisão e comentários etc.

Outros objetivos do projeto incluem a experimentação de formas colaborativas de comunicação científica e a investigação dos possíveis atores e papéis que poderiam ser envolvidos na reengenharia do processo de publicação.

Para atingir o primeiro objetivo, foi idealizado no ambiente CEJ um componente gerenciador de anotações, e o capítulo 3 desta monografia terá como objetivo traçar uma análise comparativa dos principais sistemas de anotações na Web descritos na literatura, com o objetivo de identificar as melhores oportunidades para este módulo de anotações. Um dos alicerces deste módulo gerenciador de anotações, que será vislumbrado nos capítulos seguintes, é o suporte a anotações localizadas, isto é, a possibilidade de incrementar um documento científico com anotações e linhas de discussão sobre elementos de granularidade fina do documento, como seções, parágrafos, trechos de parágrafos etc.

De acordo com ODLYZKO (2002), a necessidade de manter a revisão pelos pares em sua encarnação tradicional tem sido superdimensionada. Diversos artigos na literatura que abordam este tema da reengenharia do processo de publicação científica têm apontado a inadequação das formas convencionais de *peer-review*, e como a Internet abre um leque de possibilidades oportunas, como a revisão aberta, o uso de comentários pelos pares (*peer-commentary*) etc. Uma idéia defendida veementemente neste trabalho é que a revisão dos artigos pelos pares não produza uma decisão binária, realizada uma única vez e de forma fechada, como acontece no modelo atual. Alguns proponentes das novas formas, como Andrew Odlyzko e Stevan Harnad, têm insistido na necessidade de uma revisão pelos pares de forma *continuada*, incluindo vários ciclos de revisão, inclusive com o processo continuando depois da publicação. Neste modelo, a revisão pelos pares passa a ser considerada uma atividade “guarda-chuva”, acompanhando todo o processo de publicação, desde a definição de uma proposta de trabalho, até o período pós-publicação. Neste contexto, os artefatos gerados pela atividade de revisão podem ser considerados parte da publicação, oferecendo um feedback continuado aos leitores e pesquisadores. Esta discussão gerada pode ainda direcionar outras pesquisas na mesma linha de um artigo, agilizando o processo de comunicação científica como um todo. Para suportar a revisão de forma continuada, parece evidente este requisito de que o ambiente proposto ofereça uma

série de facilidades que apoiem a discussão, como linhas de discussão, anotações localizadas, anotações do grupo, comentários revisados etc. A revisão pelos pares poderá finalmente abandonar sua encarnação tradicionalmente fechada, tornando-se um processo aberto e colaborativo, em que os revisores fazem críticas e respondem a críticas de outros, debatem pontos polêmicos do artigo, delegam tarefas a outros revisores, convocam autores a responderem aos seus comentários, e envolvem os leitores no processo. Finalmente, a percepção do andamento do processo de publicação poderá ativar o espírito de cooperação, tornando a atividade como um todo mais ágil.

Por último, a automatização das atividades envolvidas no processo de publicação deve ser institucionalizada: um objetivo não menos importante embutido na proposta do ambiente CEJ. Esta é uma forma de reduzir os custos potenciais do processo de publicação, viabilizando um modelo econômico que tenha custo zero para os leitores, um requisito importante para o sucesso das novas formas de publicação. Em KENNEDY (2003), diversas oportunidades de automatização do processo de publicação são enumeradas, acompanhando cada uma de suas fases.

1.3 Contexto da tese

O ambiente CEJ inicialmente esteve inserido no núcleo de pesquisa DESENSUS, projeto desenvolvido na COPPE/UFRJ como uma plataforma para a geração e execução de modelos integrados de pesquisa em desenvolvimento sustentável. No início do ano de 2004, o trabalho de pesquisa no ambiente CEJ foi aceito e publicado nos anais da 8ª edição da conferência CSCWD (*Computer Supported Cooperative Work in Design*), realizada em Xiamen, na China. Também no ano de 2004, para demonstrar a viabilidade do ambiente e dos conceitos ali embutidos, alguns experimentos foram realizados com este ambiente: dois modelos inovadores de publicação científica eletrônica propostos nesta monografia foram configurados e simulados, e três experimentos reais foram realizados: dois deles com turmas de graduação da UFRJ, e o terceiro com o objetivo de revisar esta própria monografia! Estes experimentos serão descritos e analisados com detalhe no capítulo 5 desta monografia.

1.4 Organização do trabalho de tese

Esta dissertação encontra-se organizada em sete capítulos, contando com esta introdução. O capítulo 2, dedicado às infra-estruturas de conhecimento que têm emergido nos últimos anos, começa descrevendo a evolução do processo de publicação científica, desde seu advento até os dias de hoje. É realizada uma revisão bibliográfica de alguns modelos propostos recentemente, alguns inclusive implementados com sucesso em diferentes disciplinas, como alternativa ao processo tradicional de publicação constituído ao longo de várias décadas. Conforme foi dito, um dos objetivos desta análise é a identificação dos problemas e/ou limitações do processo corrente de publicação científica que poderiam ser alvo de um trabalho de tese.

No capítulo 3, é definido e caracterizado um Sistema de Anotação na Web, que será um dos alicerces fundamentais do ambiente proposto nesta monografia.

No capítulo 4, a partir dos estudos feitos abordando os problemas centrais do processo tradicional de publicação científica (no capítulos 2), do levantamento dos pontos positivos e negativos embutidos nos modelos emergentes (no capítulo 2), da caracterização do projeto de uma infra-estrutura de conhecimento (no capítulo 2) e do aprendizado adquirido no estudo dos Sistema de Anotações na Web (realizado no capítulo 3), são propostos dois modelos simples que poderiam resolver uma gama de problemas inerentes ao modelo tradicional, e é descrita a proposta do ambiente CEJ como forma de implementar estes modelos. Com o objetivo de experimentar e validar este ambiente, no capítulo 5 são apresentados alguns estudos de caso realizados no ano de 2004 com uma versão experimental do ambiente CEJ.

Finalmente, o capítulo 6 apresenta as conclusões deste trabalho, ressaltando os pontos positivos alcançados e, eventualmente, as partes ainda incompletas, incluindo recomendações para futuros trabalhos de complementação.

Capítulo 2

Infra-estruturas do conhecimento científico

2.1 Considerações iniciais

A invenção da escrita, por volta do ano de 3500 A.C, seguida pela invenção do papel por Ts'ai Lun, em 100 D.C na China, revolucionaram a transmissão e o armazenamento do conhecimento. A próxima revolução da tecnologia da informação veio por volta de 1400 D.C, quando Gutenberg inventou a prensa tipográfica. A impressão reduziu significativamente os custos e o tempo necessário para reproduzir os manuscritos.

Nas últimas décadas, com a difusão da mídia eletrônica e da internet, entramos na chamada *Galáxia Pós-Gutenberg*, "a quarta revolução no que se refere à produção de conhecimento" (HARNAD, 1991). Comparada com as revoluções anteriores, a revolução da computação, além de reduzir significativamente os custos de reprodução e armazenamento, oferece funcionalidades completamente novas: "A informação pode ser transmitida instantaneamente para longas distâncias; o conhecimento que é armazenado eletronicamente pode ser modificado até mesmo depois do momento em que é publicado..." (HARS, 2003).

Um número cada vez maior de portais na Web tem se dedicado à literatura científica nos últimos anos. Estes portais têm sido chamados de *infra-estruturas de conhecimento científico* na literatura. Apesar da rápida transição para as novas formas e o aparente progresso devido à mídia eletrônica e à Internet, uma avaliação mais cuidadosa desses portais indicaria que o modelo de publicação científica ali implementado permanece essencialmente inalterado.

Este capítulo faz uma revisão bibliográfica sobre a evolução das redes de conhecimento, começando pelas tradicionais e chegando nas emergentes, acompanhada de uma análise teórica do processo de publicação científica. O objetivo desta análise teórica é entender o sistema de publicação tradicional, seus problemas e limitações. Por exemplo, o fato de que os artigos são publicados em edições ("issues") é em função dos custos da publicação baseada em Gutenberg.

Neste sentido, a seção 2.2 aborda o processo tradicional de publicação científica e a evolução do processo de publicação, analisando-se o momento atual de transição

para as novas formas de publicação baseadas na tecnologia da informação e da Internet. Na seção 2.3 é apresentado o conceito de uma infra-estrutura de conhecimento científico, e é feita uma análise empírica de tais infra-estruturas, comparando-se alguns dos principais modelos de publicação adotados por revistas eletrônicas que emergiram nos últimos anos. Prosseguindo, na seção 2.4 é feita uma caracterização das infra-estruturas de conhecimento científico, tendo como objetivos obter as bases para o projeto do ambiente proposto nesta monografia. Finalmente, a seção 2.5 apresenta o sumário e conclusões deste capítulo.

2.2 O processo tradicional de publicação

No modelo tradicional, as publicações científicas são produzidas tipicamente por intermédio de uma entidade social, a *revista científica*. De acordo com HARS (2003), as revistas científicas emergiram na metade do século XVI, e acredita-se que as primeiras revistas surgiram em Paris e Londres. Atualmente, entre 80.000 e 100.000 revistas são publicadas no mundo inteiro (TENOR & KING, 2000), mas destas, apenas 24.000 são revisadas pelos pares (HARNAD, 2004).

Tradicionalmente, costuma-se atribuir dois papéis fundamentais às revistas científicas tradicionais no contexto do sistema acadêmico: 1) Servir de elo de comunicação da pesquisa científica; e 2) Validar a pesquisa científica com o objetivo de alocar recursos e empregos. O primeiro papel está sendo atualmente suplantado pelos repositórios de literatura científica espalhados pela Web. O segundo papel pode perfeitamente ser assumido pelos próprios pesquisadores e acadêmicos de cada área. Desta forma, o único papel exercido pelas revistas atualmente é o de fornecer rendimentos e lucros aos publicadores (BOYCE, 2000).

O objetivo desta seção é realizar uma análise teórica do processo tradicional de publicação focado nas revistas científicas. Esta análise será realizada sob três perspectivas: 1) a perspectiva do processo de publicação, que descreverá as atividades envolvidas neste processo; 2) a perspectiva dos objetos envolvidos, que descreverá as principais entidades envolvidas neste processo e suas associações; e 3) a perspectiva organizacional, que apresentará os principais atores do processo e seus objetivos.

Uma seção será também dedicada à revisão pelos pares, considerada um dos pilares de sustentação do modelo tradicional de publicação científica. Finalmente, discutiremos os principais problemas presentes no modelo tradicional.

2.2.1 Descrição do processo

O processo tradicional inicia-se com a confecção de um artigo, resultado de uma pesquisa científica que pode ser conduzida por um ou mais pesquisadores. Uma cópia do artigo é então submetida à avaliação de uma determinada revista científica. Uma revista é composta por um *corpo editorial*, que inclui um ou mais *editores* e uma equipe de *consultores*. O processo de publicação tradicional é ilustrado na figura 2.1:

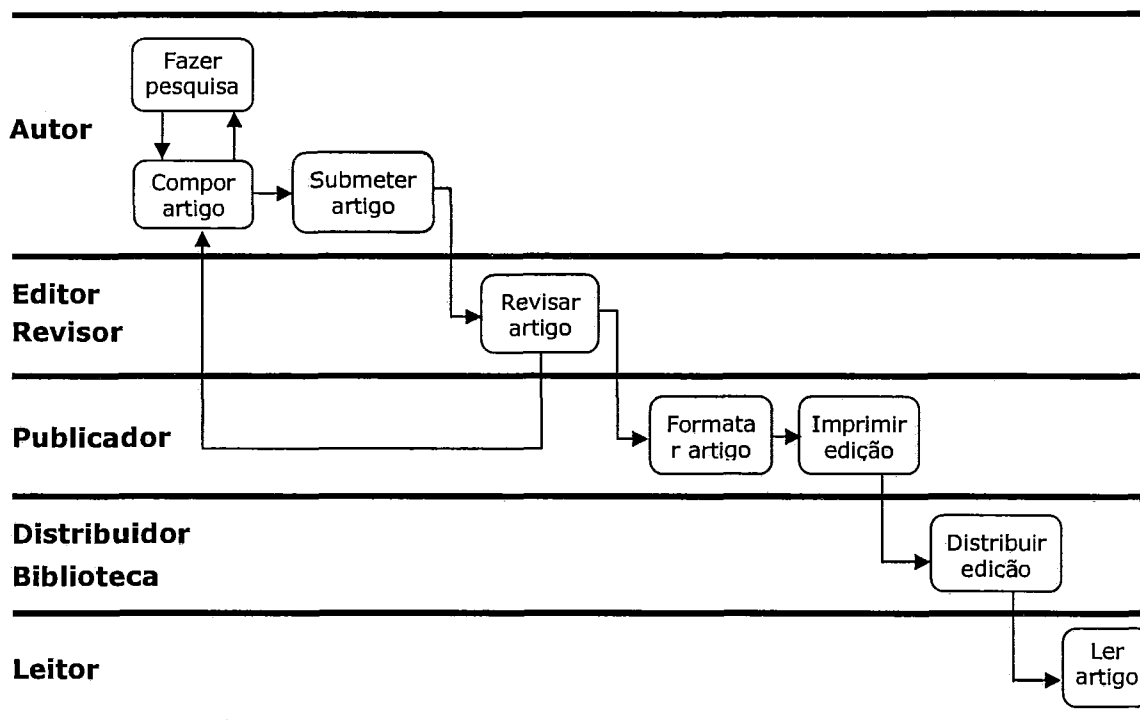


Figura 2.1 – O modelo tradicional de publicação científica

A atividade seguinte consiste na revisão ou avaliação do artigo. Para avaliar um artigo, o corpo editorial da revista geralmente seleciona dois ou três revisores. Normalmente os revisores são escolhidos de uma lista de especialistas de conhecimento dos editores, ou ainda do próprio corpo editorial. Os revisores “aconselham” os editores enviando relatórios de avaliação do artigo e fazendo recomendações acerca da aceitação ou rejeição do manuscrito. A palavra final com relação à aprovação ou não do artigo quase sempre é do editor, de forma que os revisores têm um papel de *recomendação*. O processo de revisão geralmente é realizado pela própria comunidade dos autores dos

artigos – que não recebem para isto. Por isso o processo é conhecido como “revisão pelos pares” ou *peer-review*.

Depois da revisão, caso o artigo não seja rejeitado, poderá ser necessária a edição e atualização do artigo por parte dos autores, para atender a solicitações de correções ou aprimoramentos exigidos pela equipe de revisores. Uma nova versão do artigo é então submetida e repassada aos revisores, e o ciclo continua até que o artigo seja aceito, rejeitado ou que o autor desista. Se o artigo for aceito, então é submetido à cópia-edição, que antecede a impressão. É importante destacar que o artigo somente é impresso, no modelo tradicional, quando todos os artigos de uma edição (“*issue*”) são aceitos e passam a etapa de cópia-edição. Finalmente, a edição da revista é distribuída às bibliotecas e assinantes da revista, e os artigos são disponibilizados para leitura.

Na prática, existem algumas pequenas variações deste modelo, mas esta descrição representa a essência do modelo tradicional de publicação.

2.2.2 Descrição dos objetos envolvidos

As principais entidades ou classes de objetos envolvidos no sistema de publicação tradicional são o *artigo* e a *edição* (figura 2.2). Cada artigo é associado a diversas cópias de *artigos impressos*. Uma edição representa uma coleção de artigos, e esta associação é representada pelo símbolo de agregação no diagrama de classes a seguir:

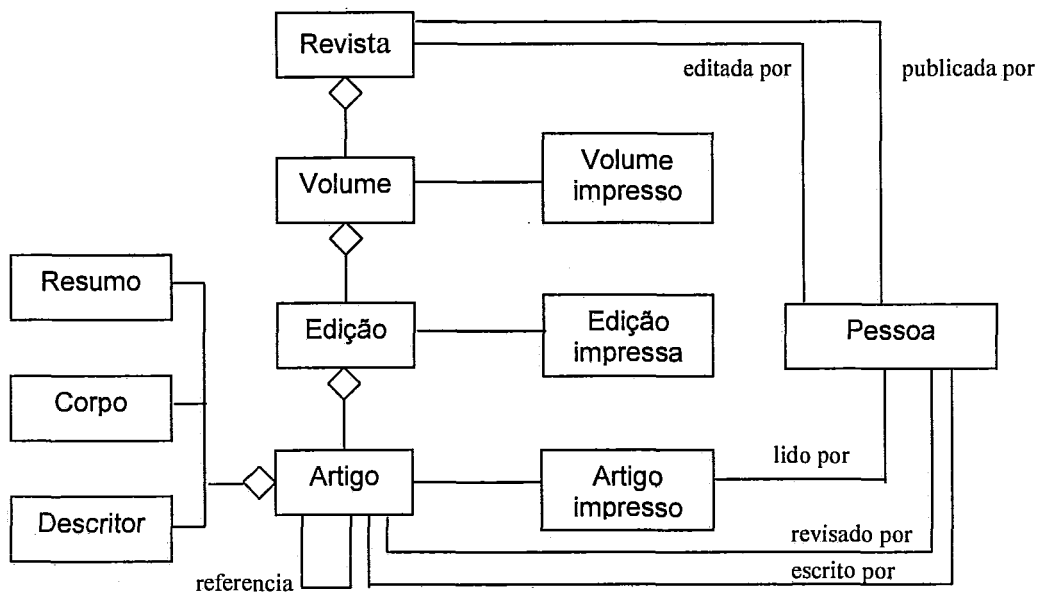


Figura 2.2 – Diagrama de classes de uma revista científica tradicional

Um conjunto de *edições* pode produzir um *volume* da revista, que pode na verdade ser vista como uma coleção de volumes. Uma revista é editada e publicada por *pessoas*. As edições e os volumes também são impressos, gerando uma *edição impressa* e um *volume impresso*. Um artigo é escrito e revisado por pesquisadores. Finalmente, o artigo é composto de um *resumo*, um *corpo* e um conjunto de *descritores*, e pode referenciar outros artigos.

2.2.3 Descrição dos aspectos organizacionais

De acordo com HARS (2003), há sete principais categorias de atores que participam no processo de publicação científica: autores, editores, revisores, publicadores, distribuidores, bibliotecas de pesquisa e leitores. Estes papéis não são exclusivos, já que um indivíduo pode assumir diversos papéis.

Os *autores* possuem diversos objetivos no sistema de publicação. De acordo com LAFOLLETTE (1992), como pesquisadores, eles objetivam difundir suas idéias entre outros pesquisadores e o avanço do conhecimento. Outros objetivos incluem o reconhecimento dos autores e recompensas acadêmicas na forma de titulações e cargos (MCKNIGHT, DILLON & SHACKEL, 1996). Para HAMILTON (1990), a publicação científica não mais representa uma forma de comunicação entre os pares, mas uma maneira de o pesquisador aumentar seu *status* e acumular pontos para a aquisição de cargos e bolsas de pesquisa.

Os *editores* são considerados como os “guardadores da ciência” (SIMON & FYFE, 1994) e objetivam aumentar a qualidade dos artigos publicados na revista e o tamanho de sua audiência. Assim como os autores, também anseiam por reconhecimento em sua área de atuação e por prêmios acadêmicos.

Por outro lado, uma análise dos objetivos dos *revisores* torna-se mais difícil, já que eles não são compensados por seu trabalho, não recebem prêmios por seu trabalho, e geralmente mantêm-se no anonimato. GLOGOFF (1988) sugere que os revisores objetivam o acesso pioneiro aos avanços em uma disciplina do conhecimento. Outros objetivos possíveis poderiam incluir reconhecimento, altruísmo, e o interesse pessoal de levantar barreiras contra seus pares (WILSON, 1978).

Já os publicadores estão interessados no sucesso da revista como um negócio, e objetivam maximizar seus lucros. De forma semelhante, os distribuidores também estão

interessados nos aspectos econômicos de suas atividades. Por outro lado, as bibliotecas de pesquisa objetivam preservar e organizar o conhecimento (OKERSON, 1996).

Finalmente, leitores buscam manter-se atualizados em seus campos de pesquisa, e muitos recorrem às revistas para assimilar conhecimento.

É importante destacar que os objetivos dos atores de um sistema de publicação frequentemente entram em conflito. No sistema atual, um pesquisador não tem condições de difundir suas idéias sozinho, em função dos altos custos da impressão e da distribuição. O objetivo dos editores de aumentar a qualidade dos artigos publicados em uma revista pode limitar a geração de idéias novas, um dos objetivos dos autores. Alguns argumentam que o sistema de publicação atual não é amigável ao leitor, em razão da organização das revistas em edições impressas, uma forma encontrada pelos publicadores para maximizar seus lucros.

2.2.4 A revisão pelos pares

O *peer-review* foi criado inicialmente como um sistema para garantir a qualidade da pesquisa publicada, mas ao longo dos anos passou também a ser usado pelas universidades como critério de escolha na promoção dos pesquisadores. Ele tem sido utilizado tanto pelas grandes revistas científicas para o controle de qualidade das publicações quanto na seleção de artigos submetidos a simpósios e conferências, ou ainda na publicação de monografias.

O processo de revisão tradicionalmente é realizado de forma *fechada* ou *anônima*, isto é, os autores do artigo não conhecem a identidade dos revisores, muito embora os revisores conheçam os nomes dos autores. Esta forma de revisão é conhecida como *closed peer-review*. Uma forma alternativa experimentada atualmente por algumas revistas é a *aberta*, em que os autores do artigo conhecem os revisores e vice-versa. Fala-se nesse caso em *open peer-review*.

Existem relatos ainda de uma terceira forma de revisão, a **duplamente fechada** (*double-blind peer-review*), em que os autores e os revisores não se conhecem reciprocamente. De acordo com CECI & PETERS (1984), essa forma de revisão tende a não funcionar porque a autoria de um artigo em geral é deduzida do contexto pelos conhecedores da área (por exemplo, através da lista de referências).

A principal crítica contra o sistema fechado está no aspecto de justiça, uma vez que os revisores, responsáveis por julgar o trabalho dos outros, o fazem de forma secreta, sem que os autores e o público leitor em geral possam conhecer sua identidade e com eles argumentar. Desta maneira, o sistema aberto tem o potencial de reduzir os abusos nas revisões, conforme será visto na próxima seção.

Uma outra vantagem do sistema aberto bastante defendida é o fato de ela possibilitar o reconhecimento também dos revisores, e não apenas dos autores. Ocasionalmente os revisores podem contribuir com o trabalho de outros pesquisadores, e o sistema aberto ainda evita o roubo de idéias de outros autores e pesquisadores.

O principal argumento contra o sistema de revisão aberta é que revisores juniores serão relutantes em criticar os trabalhos de pesquisadores seniores por medo de represálias (SMITH, 1999a).

2.2.5 Problemas com o processo de publicação científica

2.2.5.1 Lentidão do processo

Um dos problemas inerentes ao processo de publicação científica tradicional, baseado fundamentalmente na mídia de Gutenberg – o papel impresso - está na *velocidade* com que são produzidas as publicações. O intervalo entre as datas de submissão e de publicação de um artigo geralmente é muito longo, podendo chegar a 2 ou 3 anos em alguns casos. Isto se deve em grande parte à organização das revistas científicas em edições. No modelo tradicional de publicação, o objeto básico que é impresso e distribuído não é um artigo, mas uma coleção de artigos – a edição ou “*issue*”. Isto acontece em função dos aspectos econômicos envolvidos: existe um custo fixo elevado para distribuir e imprimir uma coleção de artigos, tornando mais vantajoso distribuir a coleção inteira ao invés de artigos individualmente. Deste modo, o fato de um artigo já ter sido aprovado pela revisão pelos pares não implica em publicação imediata: deve-se esperar a publicação de uma “*issue*”.

Por outro lado, o processo tradicional de publicação é seqüencial, onde a revisão ocorre sempre antes da impressão, que acontece antes da distribuição, que acontece antes da leitura.

2.2.5.2 Crise dos seriais

Um outro problema do modelo atual é o aumento exponencial do número de artigos publicados em revistas científicas de uma determinada área ou disciplina, que ultrapassou a capacidade de leitura e absorção de um pesquisador.

Com a estagnação dos orçamentos das bibliotecas – universitárias e especializadas – para aquisição de informação, os editores de publicações científicas passaram a ter que competir por recursos cada vez mais escassos em orçamentos reduzidos (ODLYZKO, 1995). O fato é que o sistema atual de publicação, constituído ao longo de diversas décadas, se tornou extremamente caro para ser sustentado pela comunidade acadêmica, em função de uma série de fatores, entre os quais destaca-se o aumento exagerado do volume e dos custos das publicações. Este fenômeno ficou conhecido na literatura como a *crise dos seriais*.

Diversos estudos e conferências têm procurado analisar as causas dessa crise, e apontado possíveis soluções para o problema. Nos últimos anos, a publicação eletrônica começa a aparecer como tema de diversos seminários e conferências internacionais.

2.2.5.3 Revisão pelos pares fechada, binária e uma única vez

O processo de revisão pelos pares também tem sido alvo de inúmeras críticas na literatura dedicada ao assunto. Este processo produz como resultado final uma decisão binária de publicar ou não o artigo, tomada uma única vez, e adequada apenas à validação da pesquisa para a alocação de empregos e cargos. Não oferece nenhum benefício aos leitores e é muito limitada aos autores. Não prestigia os revisores envolvidos no trabalho de avaliação, em função da abordagem anônima ou fechada de avaliação.

HARNAD (1998) enumera diversos problemas do modelo corrente de avaliação da literatura realizada pelas revistas:

- Os editores das grandes revistas freqüentemente se inclinam a um determinado avaliador ou outro, beneficiando uns em detrimento de outros;
- Alguns avaliadores podem não ser especialistas no assunto, mal-informados ou ainda injustos na avaliação;

- Os autores também têm sua parcela de culpa. Eles acabam publicando seus trabalhos em revistas menos criteriosas, se as mais rigorosas rejeitarem seus artigos. Existe uma hierarquia de revistas, indo das mais rigorosas até as mais fracas.

WILLIAMSON (2002) divide os problemas relacionados ao *peer-review* nas seguintes categorias: subjetividade, tendenciosidade, abuso pelos autores, abuso pelos revisores, falhas na detecção de erros, fraude e má conduta.

Problemas de *subjetividade* incluem o caso de artigos que não chegam a ser enviados aos revisores porque têm seus resumos rejeitados, e editores que escolhem determinados revisores quando querem ver um artigo rejeitado. *Tendenciosidade* se refere à discriminação contra autores em função de sua nacionalidade, linguagem nativa, sexo ou instituição, ou ainda à situação em que os revisores competem de alguma maneira com o autor do artigo submetido à publicação.

Abuso pelos autores ocorrem quando publicam demasiadamente artigos em uma área de pesquisa (“salami”), quando fazem “clonagem” de outras publicações, ou ainda quando omitem ou degradam o trabalho de pesquisadores a eles subordinados.

Abuso pelos revisores também são comuns, e acontecem geralmente quando estes publicam o trabalho de outros (pré-publicação) que foi enviado para revisão (*plagiarismo*), ou ainda quando atrasam a publicação de um trabalho que de alguma maneira compete com outro de interesse do editor ou revisor.

É comum também existirem *falhas na detecção de erros* nos artigos por parte dos revisores, e finalmente, problemas de *fraude e má conduta*, quando os autores fabricam resultados, falsificam dados ou reivindicam a autoria de resultados de outros.

Em muitas disciplinas faltam incentivos para os autores pré-publicarem seus artigos na Web. Se eles o fazem, limitam suas opções de publicação em revistas especializadas em função de elas obrigarem o autor a transferir seus direitos a elas. Estas revistas dificultam a disseminação ampla da pesquisa através de taxas de inscrição para acesso à pesquisa e de políticas que restringem a publicação de pré-impressões e a disseminação de pós-publicações. Chega-se a um impasse: se o autor decide publicar seu artigo numa revista especializada que impede a publicação livre na Web do artigo, o autor não terá seu trabalho disponível livremente na Web. Assim, seu artigo será menos citado e, portanto, terá menos valor. Por outro lado, se ele decide publicar na Web seu trabalho – na sua página pessoal, por exemplo – muitas revistas de nome não

aceitarão seu artigo para publicação, já que o autor estaria burlando uma das normas de aceitação de trabalhos.

2.3 Transição para as novas formas de publicação

O advento da era da informação e do ambiente eletrônico, considerado em conjunto com os problemas discutidos na seção anterior, trouxe à tona a necessidade de se repensar o processo de revisão da literatura à luz das novas tecnologias. O momento atual é de experimentação, em que revistas emergentes propõem novos modelos para o processo de publicação científica, buscando resolver ou minimizar alguns dos problemas apresentados pelo modelo corrente.

Como vimos, é urgente que novos modelos para a comunicação científica sejam propostos e discutidos, como forma de minimizar os problemas e impasses que apresenta no modelo corrente. Primeiro, é preciso impedir que a mídia eletrônica seja utilizada para "clonar" o formato tradicional da literatura científica baseado no papel, como vem acontecendo. É necessário quebrar o ciclo envolvendo apenas as grandes revistas e as bibliotecas de pesquisa. Cabe aos pesquisadores, que são os produtores e consumidores da informação, a tarefa de tomar a frente nesse processo.

Segundo, a reforma do *peer-review* precisa ser encaminhada. As tentativas de conversão das revistas tradicionais para o formato eletrônico na grande rede simplesmente mantiveram os mesmos mecanismos de validação do modelo antigo, não questionando se esses mecanismos eram bons na nova realidade. Desta forma, nas infraestruturas atuais, o processo de revisão continua produzindo uma decisão única e binária (ODLYZKO, 1995). Em SHUM & SUMNER (1996), por exemplo, a atividade de revisão pelos pares é analisada sobre a perspectiva do que eles chamam de "argumentação colaborativa suportada por computador" ou, em inglês, "*Computer-Supported Collaborative Argumentation (CSCA)*". Os autores propõem um novo modelo para a validação da literatura pelos pares, que é vista como um processo argumentativo. Desta forma, são apontadas algumas limitações adicionais do modelo corrente, como o fato de que perguntas freqüentemente ficam sem resposta, e contradições muitas vezes não são devidamente elucidadas durante a atividade de revisão. Isto em função de a interatividade presente no debate face-a-face típico de conferências não existir. "A revisão (no modelo atual) é um processo argumentativo onde revisores se engajam em um debate imaginário com autores distantes, que não estão presentes para responder às suas análises" (SHUM & SUMNER, 1996).

2.3.1 A mídia eletrônica

De acordo com ODLYZKO (1994), o papel é uma mídia “difícil de produzir, difícil de distribuir, difícil de arquivar, e difícil de duplicar”. A mídia eletrônica, por outro lado, não apresenta nenhuma das características anteriores, favorecendo a disseminação do conhecimento em larga escala de forma eficiente e a baixo custo.

O problema é que a mídia eletrônica vem sendo utilizada para encarnar um "clone eletrônico" da versão em papel das grandes revistas, como temos comentado. Estas têm insistido na preservação dos principais mecanismos do modelo antigo.

Um exemplo que ilustra o que dissemos é a escolha pelo padrão *Portable Document Format* (PDF), um formato proprietário que é utilizado pela maioria das revistas tradicionais, e mesmo as que possuem suas versões na Web, para a confecção e distribuição dos artigos científicos. O formato PDF, sendo orientado a páginas, é muito útil na impressão de documentos científicos, e é utilizado extensivamente na confecção de revistas e anais de conferências. Todavia, esta tecnologia apresenta uma série de limitações quando o objetivo é manipular com o documento, por exemplo, quando queremos acrescentar anotações ou linhas de discussão sobre um determinado documento. Mais grave ainda é o fato de que não se compatibiliza com a infra-estrutura da Web, não sendo possível interagir com um documento PDF da mesma forma como é possível manipular com documentos HTML ou XML através do modelo de objetos do documento – *Document Object Model* (DOM).

Nos últimos anos, a publicação eletrônica começa a aparecer como tema de diversos seminários e conferências internacionais, em que os principais participantes têm sido os editores das grandes revistas e membros ligados às bibliotecas de pesquisa. A própria Adobe, empresa criadora do formato PDF, tem promovido conferências anuais, em conjunto com as grandes revistas, com o objetivo de difundir novas ferramentas e funcionalidades associadas à tecnologia PDF. Os pesquisadores, principais atores desse processo (atuando como autores, leitores e revisores), surpreendentemente não têm sido atuantes. Na discussão de um novo modelo, os pesquisadores deveriam constituir a voz mais alta, já que são os principais interessados no processo ao atuarem como produtores e consumidores da informação.

2.3.2 O papel dos repositórios eletrônicos (*eprints*)

Com o advento da Web, popularizaram-se na última década os chamados repositórios eletrônicos de literatura científica. Estes repositórios são servidores de arquivo que disponibilizam para o público em geral manuscritos de artigos de diversas disciplinas do conhecimento, tanto revisados quanto não revisados.

Estes repositórios nasceram inicialmente em algumas áreas da Física, sobretudo a Física de Alta Energia. De acordo com HARNAD (2003), um dos mais proeminentes defensores dos repositórios abertos de literatura científica, existem evidências de que a Física possui não apenas uma tradição no compartilhamento de manuscritos de artigos, como apresenta taxas de rejeição de artigos (pelas revistas) menores do que em outras disciplinas. Não que o trabalho de pesquisa dos físicos seja de qualidade inferior. Harnad defende que os físicos possuem uma cultura mais "realista" a respeito da qualidade de seus trabalhos, submetendo-os diretamente às revistas apropriadas (já que existe uma hierarquia evidente de qualidade e rigor entre as revistas científicas). Desta forma, em muitas áreas da Física, os manuscritos antes e após a revisão podem ser muito próximos, novamente porque os autores os prepararam mais realisticamente, sem esperar que os revisores apontassem eventuais falhas.

LUCE (2000) ilustra como os repositórios de literatura científica estão evoluindo e cita alguns desafios que se impõem e precisam ser atacados:

- A integração dos repositórios eletrônicos ao sistema de informação acadêmico precisa ser institucionalizada.
- Melhores recursos na busca e na recuperação da literatura, especialmente entre disciplinas, será fundamental para o trabalho do pesquisador.
- Desenvolver um sistema com interfaces amigáveis ao usuário é imperativo para o sucesso dos repositórios eletrônicos.
- Toda gama de informação relevante, incluindo metadados, comentários, texto completo e citações, deve ser incluída nos repositórios.

Ainda de acordo com HARNAD (2003), o acesso gratuito e online ao resultado de toda a pesquisa científica revisada, através dos repositórios abertos, é uma solução ótima e ao mesmo tempo inevitável. No dia em que os dois milhões de artigos publicados anualmente em 24000 revistas especializadas estiverem acessíveis na Web sem restrições de acesso, ao público em geral, a pesquisa científica será muito mais rápida e interativa. Cada artigo poderá ser ligado diretamente a todo artigo referenciado;

novas formas de revisão como os comentários pelos pares irão enriquecer a pesquisa científica; ciclos regenerativos de interação poderão agilizar sobremaneira o processo de validação; e finalmente, novas ferramentas irão oferecer medidas mais precisas e realistas do uso e do impacto da pesquisa científica, como já vem sendo feito pelas máquinas de busca citebase <<http://citebase.eprints.org>> e citeseer <<http://citeseer.nj.nec.com/cs>>.

2.3.3 O impacto da tecnologia da informação

A tecnologia da informação abre um leque de possibilidades inimagináveis de melhorias no processo de comunicação acadêmica. A análise teórica realizada nas seções anteriores pode ser utilizada para uma análise destas eventuais modificações.

HARS (2003) estabelece um conjunto de 11 proposições com o objetivo de traçar as principais diferenças entre as revistas eletrônicas e as revistas tradicionais, antecipando as características fundamentais das novas revistas. Nesta seção faremos uma análise resumida de cada uma delas.

Proposição #1 – Documentos eletrônicos podem ser modificados depois de publicados

Uma característica importante que distingue os documentos eletrônicos dos documentos escritos em papel é a habilidade de atualizar seu conteúdo. Isto é consequência da tecnologia dos discos rígidos que permite apagar e reescrever a informação repetidas vezes. Desta forma, tecnicamente é viável atualizar as publicações eletrônicas disponíveis na Web com o objetivo de clarificar alguns pontos, incluir sugestões de revisores e leitores, corrigir erros ou acrescentar novos relacionamentos identificados. Todavia, o fato de isto ser viável não implica em sua adoção pelas revistas eletrônicas. A maioria das revistas atuais não permite que os autores modifiquem o conteúdo de seus artigos depois de publicados, embora existam algumas exceções (por exemplo, a revista *Communications of the AIS*).

Segundo HARS (2003), a possibilidade de alterar o conteúdo dos artigos após sua publicação irá aumentar, e não reduzir, a qualidade dos artigos. Além disso, os leitores terão a possibilidade de escolher que versão desejam ler: uma versão anterior revisada, ou versões subsequentes não-revisadas porém aprimoradas. Seguindo esta linha de raciocínio, o autor sugere que as revistas eletrônicas que desejarem maximizar

o valor dos artigos nela publicados irão permitir que autores modifiquem seus documentos. Isto leva à proposição 2:

Proposição #2 – Artigos poderão ser atualizados pelos próprios autores

Uma outra consequência da proposição 1 é que, em contraste com a mídia impressa, **em uma revista eletrônica, a publicação de um artigo é reversível!**

Proposição #3 – Cada leitor poderá visualizar um documento de uma forma diferente

Comparando-se mais uma vez a mídia impressa com a eletrônica, documentos eletrônicos podem ser apresentados ao leitor de modos diferentes. De acordo com as necessidades e o nível de especialização do leitor, algumas informações do artigo poderiam estar visíveis ou não; dados poderiam ser exibidos de diferentes maneiras; fórmulas matemáticas poderiam ser demonstradas *online*, tornando o artigo interativo.

Proposição #4 – Na Web, a localização física de um artigo e de um leitor é irrelevante

Graças à revolução da Internet, o acesso a um documento eletrônico ou a qualquer outro tipo de informação é realizado de forma quase instantânea independente do local do planeta, e de forma transparente ao leitor. O acesso a documentos através de uma URL (Uniform Resource Locator) permite que arquivos locais e externos sejam tratados da mesma maneira. De forma que, na Web, a localização física de um artigo, da mesma forma que o leitor, é irrelevante. É verdade que a independência da localização não é atingida completamente, já que os tempos de resposta ainda dependem da distância do leitor do artigo, mas isto tende a diminuir drasticamente em função da evolução da Internet, sobretudo com a popularização da banda larga.

Uma consequência da proposição 4 é que, **se um documento consiste de diversas partes, é irrelevante onde cada parte está localizada.** Para o leitor, se um documento é composto de diversas seções, é irrelevante se algumas seções são recuperadas localmente ou de um país distante.

Proposição #5 – Numa revista eletrônica, todos os leitores compartilham um mesmo documento (não existem cópias)

Já que a localização de um artigo é irrelevante na Internet, de acordo com a proposição 4, não há necessidade de armazenar cópias de um artigo localmente ou em bibliotecas para simplificar o acesso. A tecnologia da informação permite o acesso simultâneo e paralelo ao mesmo arquivo físico – original – a diversos leitores, de forma transparente. Na prática, quando visualizamos um documento na Web, uma “cópia” é feita e armazenada no cachê do usuário. Mas esta cópia é temporária, ficando ali guardada por um curto espaço de tempo.

Uma importante diferença do mecanismo de recuperação de um documento eletrônico em relação ao processo de cópia de um artigo impresso é que, no ambiente eletrônico, é possível que somente as partes do documento que estão sendo lidas sejam recuperadas. Isto tudo nos leva a proposição 5: numa revista eletrônica, todos os leitores compartilham um mesmo documento, e não há necessidade de manter cópias.

Uma outra diferença entre os ambientes eletrônico e impresso é que, no impresso, anotações sobre o documento somente podem ser feitas localmente, sobre uma cópia do documento. A Internet, por outro lado, traz o potencial de centralizar todas as anotações de um documento em um único servidor, sem necessariamente exigir que as anotações sejam todas públicas. Neste sentido, uma consequência imediata da proposição 5 é que, **na Internet, não há necessidade de manter cópias de um documento público: apontadores (*links*) para o original são preferíveis.**

Proposição #6 – Revistas não serão mais organizadas por edições (*issues*)

A emergência dos documentos eletrônicos não implica automaticamente que os documentos impressos irão tornar-se obsoletos. Muitas revistas eletrônicas são híbridas, isto é, publicam tanto versões impressas quanto online de suas publicações. Muitos leitores ainda preferem a leitura em papel à leitura sobre o display de um monitor: para muitos leitores, a tecnologia empregada nos monitores ainda é imatura, apesar do seu rápido aperfeiçoamento. Uma boa notícia é que os preços dos monitores de tecnologia “flat” estão caindo significativamente (HARS, 2003), sugerindo que em poucos anos sua adoção maciça seja uma realidade. Recentes pesquisas da IBM e da XEROX levaram à construção de monitores *flat* com uma resolução semelhante à do papel.

Monitores flexíveis com uma resolução de 200 DPI, que são ligeiramente mais espessos do que o papel, têm sido demonstrados nos últimos anos (MANN, 2001). Desta forma, não seria irrealista assumir que, no futuro próximo, as novas tecnologias de displays serão suficientemente maduras para substituir o papel.

A eliminação das cópias dos artigos (proposição 5) e dos documentos impressos traz importantes modificações para o processo de publicação científica. Em contraste com o ambiente impresso, onde a distribuição do artigo é um pré-requisito para sua leitura, no ambiente eletrônico a distribuição se torna um aspecto da própria leitura, acontecendo em paralelo com esta: na medida em que o leitor vai navegando sobre um documento, as partes que vão sendo lidas vão sendo aos poucos recuperadas.

No modelo tradicional de publicação, o objeto básico que é impresso e distribuído não é um artigo, mas uma coleção de artigos – a edição ou “*issue*”. Isto acontece em função dos aspectos econômicos envolvidos: existe um custo fixo elevado para distribuir e imprimir uma coleção de artigos. Por exemplo, o custo de distribuir um único artigo não é muito diferente do custo de enviar uma edição inteira (*issue*) da revista. Na Internet, todavia, custos de distribuição são medidos pela largura da banda, e não há vantagem alguma em transmitir uma coleção inteira de artigos ao invés de artigos individualmente. Ao contrário: a distribuição individual de artigos oferece um custo-benefício maior do que a distribuição de coleções inteiras de artigos (MCKNIGHT, DILLON & SHACKEL, 1996). Isto nos leva à proposição #6: revistas eletrônicas não serão mais organizadas por edições (*issues*).

Esta tendência inclusive já é visível em diversas revistas eletrônicas que praticam a chamada “pré-publicação”, onde um artigo é disponibilizado antes que a edição da revista seja liberada. Em outros casos, como acontece com a revista “Journal of Artificial Intelligence Research”, o conceito de “*issue*” foi eliminado completamente. Argumento semelhante ao utilizado para justificar o eventual desaparecimento do conceito de **edição** sugere que o conceito de **volume**, da forma como existe no modelo tradicional, também entrará em desuso. Na Web, o acesso a um artigo é realizado através de mecanismos distintos do ambiente impresso: através de poderosas máquinas de busca, via mapas por assunto ou outros esquemas de classificação que serão discutidos na seção 2.4.2 deste capítulo.

Proposição #7 – Na Web, pesquisadores podem auto-publicar seus artigos

No ambiente impresso, a edição (*copy-editing*) é um pré-requisito para publicar um artigo. Ela prepara o manuscrito para a impressão e pode também melhorar a linguagem, a compreensão, consistência e apresentação do trabalho (O'CONNOR, 1986). No ambiente eletrônico o autor já fornece o artigo no formato apropriado para distribuição ou num formato que possa ser convertido automaticamente. Desta forma, a publicação “direta”, sem necessidade de edição, é tecnicamente viável. Por outro lado, tudo que é preciso para publicar um artigo na Web é uma conexão com a Internet e um espaço em um servidor de arquivos. A maioria dos pesquisadores possui acesso a Internet, e espaço para publicar artigos não é um recurso escasso. Isto nos leva à proposição #7: na Web, pesquisadores podem auto-publicar seus artigos (ARMSTRONG, 1997). A consequência imediata desta proposição é que **a publicação deixará de ser uma função importante das revistas eletrônicas**. Se o espaço na Web para a publicação não é um recurso escasso, então uma revista eletrônica, ao oferecer um espaço físico para o armazenamento de um artigo, não adiciona nenhum valor aos autores.

Proposição #8 – Uma revista eletrônica poderá aceitar artigos submetidos por usuários que não sejam autores do artigo

Se um determinado pesquisador tiver auto-publicado um artigo seu na Web, de acordo com a proposição #7, então este artigo estará acessível pela Internet, e poderá ser incorporado em qualquer revista eletrônica através de um apontador (consequência da proposição #5) sem necessidade de autorização explícita do autor. Desta forma, ao contrário do que acontece no ambiente impresso em que somente o próprio autor pode submeter um artigo de sua autoria, no ambiente eletrônico qualquer pessoa poderá “incluir” um determinado artigo em uma revista, desde que ele esteja acessível publicamente na Web, através de uma URL. Tecnicamente é suficiente um apontador para o artigo, e isto não constitui uma violação dos direitos autorais do autor. Isto leva à proposição #8: uma revista eletrônica poderá aceitar artigos submetidos por usuários que não sejam autores do artigo.

Este fato traz ainda uma nova alternativa às revistas eletrônicas: ao invés de esperar por submissões, os editores de uma revista eletrônica poderiam vasculhar a Web à procura de documentos científicos, procurando artigos que se encaixem no escopo editorial da revista e buscando maximizar a qualidade.

Proposição #9 – Revisões serão multidimensionais, não binárias

Conforme comentado na seção 2.2.1.3 deste capítulo, o processo de revisão pelos pares, considerado um dos principais alicerces dos processos de publicação e comunicação científicas, tem sido alvo de inúmeras críticas. De acordo com HARS (2003), os cientistas cultivam uma visão cética com relação à ciência e ao conhecimento, e aceitam somente o conhecimento claramente e racionalmente demonstrado. Esta posição foi institucionalizada através do processo de revisão pelos pares, onde os artigos são revisados *antes* de sua publicação, para garantir qualidade.

A transição para as novas formas de publicação abre um leque de oportunidades de aprimoramento no processo tradicional de revisão. GINSPARG (1996) argumenta que “a implementação corrente da revisão pelos pares, da forma como é empregada pelas revistas tradicionais, precisa ser inteiramente repensada tendo em vista as novas possibilidades oferecidas pela publicação e pela distribuição eletrônicas”. VARIAN (1997) identificou brilhantemente por que a revisão pelos pares, da forma como é empregada pelas revistas científicas tradicionais, é inadequada na nova realidade:

“Quando a publicação era uma atividade cara e custosa, era apropriado avaliar artigos antes de sua publicação. Idealmente, somente manuscritos interessantes e rigorosamente carimbados poderiam prosseguir no processo de publicação. Além disso, a publicação era um sinal binário: ou bem o artigo era publicado ou bem era descartado. A publicação eletrônica, por outro lado, é barata. Essencialmente, tudo deve ser publicado, no sentido de ser disponibilizado para *download*. Mecanismos de filtro auxiliarão os leitores na tarefa de definir que artigos merecem ser baixados e lidos. O sistema atual de revisão pelos pares poderia simplesmente ser clonado no ambiente eletrônico. Mas a mídia eletrônica oferece possibilidades que não seriam possíveis na mídia impressa. Outras formas de filtro poderiam ser mais efetivas e eficientes”.

Como se pode notar nas palavras de VARIAN (1997), um dos problemas centrais do processo tradicional de revisão, já comentado anteriormente, é a natureza binária deste processo. Isto é consequência do espaço para publicação ser caro e escasso no ambiente impresso. Desta forma, muito embora as informações oriundas da revisão

de um artigo sejam extremamente úteis para o leitor, a escassez de espaço nas revistas impressas impede que este conhecimento seja publicado junto com o artigo.

No ambiente eletrônico, todos os artigos podem ser publicados, de acordo com a proposição #7, e a revisão pelos pares não precisa ser empregada para resolver conflitos acerca do que vai ou não ser publicado numa revista! O espaço para publicar é abundante, e desta forma a informação gerada pela revisão pelos pares pode ser publicada junto com o artigo. De acordo com HILGARTNER (1995), a divisão binária entre o que foi publicado e o que não foi publicado pode ser substituída por um espectro. Isto nos leva à proposição #9: no ambiente eletrônico, revisões serão multidimensionais, não binárias.

Um esquema multidimensional de avaliação tem muitas vantagens. Ele transfere a tarefa dos revisores de decidir pela aceitação ou rejeição do manuscrito para a tarefa de levantar as vantagens e problemas de um artigo (HARS, 2003). A forma como esta informação poderia ser agregada e utilizada para definir uma “aceitação” ou “rejeição” do artigo poderia ser deixada a cargo do pesquisador. Tecnicamente, esta “agregação” das informações referentes à avaliação de um artigo pode ser alcançada com a adição de funções de filtro em uma infra-estrutura de conhecimento, o que será analisado na seção 2.4.2 deste capítulo. Mecanismos de filtro podem também ser empregados permitindo que leitores acessem somente os artigos nos quais estejam interessados. A separação da revisão da decisão de publicar ou não o artigo traz a vantagem de que a revisão poderá oferecer informações detalhadas acerca das contribuições daquele artigo, o que não acontece no modelo tradicional de publicar ou não. Além disso, permite que se façam distinções entre os artigos de melhor qualidade, o que não acontece no modelo tradicional. Por exemplo, nenhuma distinção é feita entre os artigos publicados numa mesma edição das inúmeras revistas periódicas da ACM.

Proposição #10 – Artigos serão primeiro publicados, depois revisados

De acordo com HARS (2003), no modelo tradicional de publicação, a seqüência natural das fases que constituem este processo é óbvia: a revisão deve ocorrer antes da impressão, que deve acontecer antes da distribuição, que deve ocorrer antes da leitura. No ambiente eletrônico, por outro lado, a impressão é eliminada e a distribuição se torna um aspecto da leitura. Desta forma, esta seqüência se reduz a revisar → ler.

Ora, no ambiente eletrônico o espaço é abundante (proposição #7) e os artigos podem ser modificados depois de publicados (proposição #2), além de ser tecnicamente possível remover um artigo (o apontador, na verdade) de uma revista, mesmo depois de publicado. Desta forma, é possível publicar um artigo primeiro e depois revisá-lo, o que reduziria o ciclo de publicação na medida em que a revisão seria executada em paralelo com a leitura do artigo. Isto nos leva à proposição #10: no ambiente eletrônico, artigos serão primeiro publicados, depois revisados.

Isto já acontece desde 1991, quando o físico Paul Ginsparg instituiu um dos maiores repositórios de literatura científica abertos. O sucesso da iniciativa de Ginsparg levou ao desenvolvimento de outros repositórios, em outras áreas da ciência.

Proposição #11 – Um artigo pode aparecer simultaneamente em diversas revistas

Finalmente, se uma revista eletrônica pode incorporar artigos que não estejam sob seu controle físico direto, então um artigo poderá ser publicado em diversas revistas ao mesmo tempo, o que nos leva à proposição #11. Uma revista que deseja oferecer aos seus leitores artigos de excelente qualidade poderá fazê-lo mesmo que a primeira revisão sobre o artigo tenha sido feita por outra revista. Isto pode exigir em alguns casos que a revista da casa analise a confiabilidade do processo de revisor da revista concorrente que revisou o artigo. A conclusão a que chegamos é que diversas revistas poderão lucrar com o mesmo artigo, no processo de publicação redesenhado.

2.4 Infra-estruturas de conhecimento científico

Na seção anterior fizemos uma análise teórica do modelo tradicional de publicação científica centrado nas grandes revistas, e analisamos também a transição para as novas formas de publicação, decorrente do advento da mídia eletrônica e da Internet. Discutimos também o impacto da tecnologia da informação na definição de um novo modelo para as revistas científicas, apresentando as características que, de acordo com HARS (2003), possuem a maior probabilidade de definir este novo modelo.

O objetivo desta seção é apresentar as revistas eletrônicas dentro de um contexto mais geral: o das infra-estruturas de conhecimento. Faremos também uma comparação de alguns modelos de publicação adotados por revistas eletrônicas que emergiram nos últimos anos. Isto permitirá, entre outras coisas, identificar algumas características já discutidas na análise teórica e que tenham sido implementadas com sucesso.

HARS (2003) define o termo *infra-estrutura de conhecimento científico* como um “sistema de informação sócio-técnico centrado na criação, organização, armazenamento e aplicação do conhecimento científico. Não se trata apenas de um sistema técnico que consiste de hardware, software, bancos de dados etc. É também um sistema organizacional, envolvendo seus atores, estruturas organizacionais e processos”. Alguns autores preferem o termo “redes de conhecimento”. Uma grande parte desses sistemas pode ser categorizada como revistas eletrônicas. A figura 2.3 exibe uma caracterização genérica de tais infra-estruturas, adaptada também de HARS (2003):

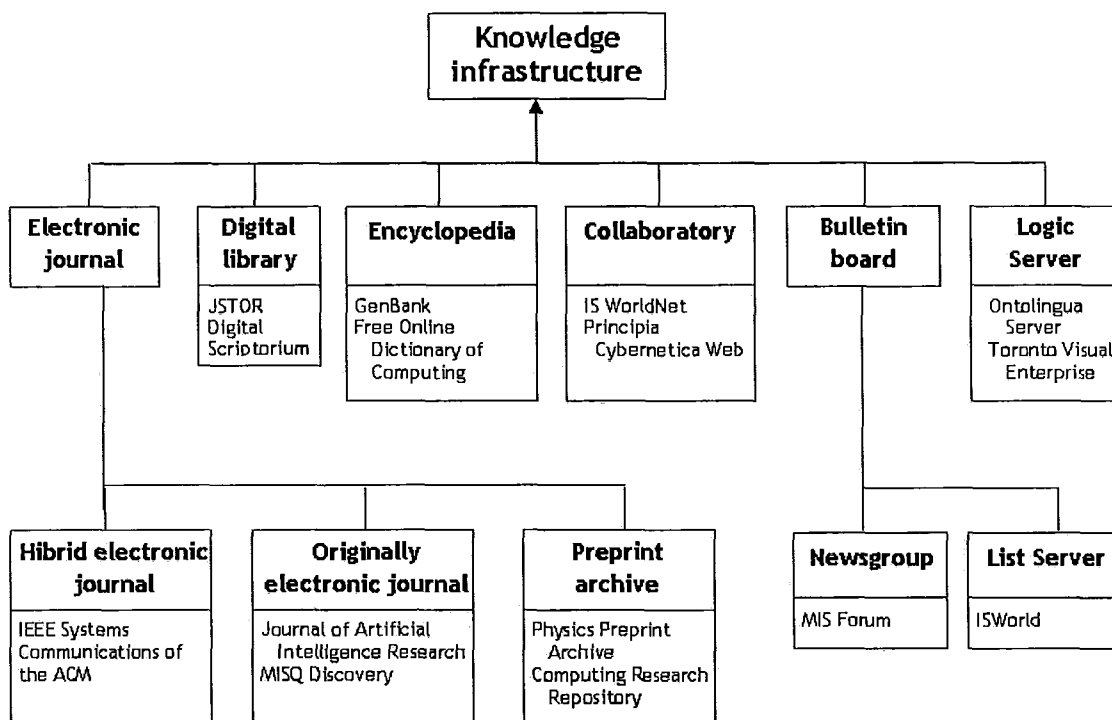


Figure 2.3 - Categorias de infra-estruturas do conhecimento (HARS, 2003)

Com a emergência de tais infra-estruturas, estamos caminhando rumo a um processo de publicação continuado, e isso exige também uma revisão continuada, que irá fornecer um *feedback* constante dos artigos aos pesquisadores, e não apenas uma decisão binária, tomada uma única vez, típica das revistas tradicionais (ODLYZKO, 2003).

Todavia, preocupações a respeito de possíveis prejuízos ao sistema tradicional de revisão pelos pares estão retardando o passo da mudança. Apesar do aparente progresso, no qual as revistas científicas tradicionais estão migrando para a Web, formas inovadoras de comunicação e publicação eletrônica estão emergindo, e algumas

atividades do processo de publicação estão sendo automatizadas, o modelo de publicação científica baseado no papel, e centrado nas grandes revistas, permanece essencialmente inalterado, como comentado anteriormente.

Por estarmos num momento de transição, em que o cenário mais provável é a derrocada do modelo tradicional, face às inúmeras contradições em sustentar-se no ambiente eletrônico, seria um risco afirmar que tais infra-estruturas seriam representativas do novo modelo. É bem verdade que algumas revistas eletrônicas vêm experimentando novas formas de discurso acadêmico, e alguns melhoramentos significativos têm sido alcançados em algumas disciplinas. Muitos pontos apresentados nas onze proposições de HARS (2003), descritos na seção anterior, de alguma forma começam a se tornar realidade nestas revistas. Este é o caso, por exemplo, da revista *BMJ*, na área de Medicina, da revista *Psychology* na área da Psicologia, e do revolucionário modelo de publicação da Física de Alta Energia. Esses e outros modelos serão analisados na seção seguinte. No entanto, se considerarmos o universo de 24.000 revistas científicas publicadas no mundo inteiro (HARNAD, 2004), esses casos constituem ainda a exceção, e não a regra.

2.4.1 Categorias de infra-estruturas de conhecimento

Considerando o conjunto de infra-estruturas de conhecimento descritas por HARS (2003), uma grande parte delas pertence à categoria de *revistas eletrônicas*. Alguns autores empregam este termo para descrever as infra-estruturas baseadas na tecnologia da informação que eventualmente substituirão as revistas tradicionais em papel (WATSON, 1994), enquanto outros o utilizam para descrever os portais que replicam as estruturas típicas do modelo tradicional de publicação.

Uma subcategoria das revistas eletrônicas são as *revistas híbridas*, que representam revistas tradicionalmente baseadas no papel que agora oferecem também versões eletrônicas de suas publicações. O acesso às publicações dessas revistas é restrito, e em alguns casos segue o modelo *pay-per-view*. Normalmente estas revistas também são organizadas por edições e volumes, e oferecem acesso gratuito aos resumos e palavras-chaves de um artigo. O texto completo dos artigos geralmente é oferecido no formato PDF e, em menor escala, nos formatos PS ou HTML. Nesta categoria podemos incluir as revistas da ACM, IEEE, Elsevier e Springer. Não é difícil concluir que as

revistas híbridas atuais constituem, na verdade, clones eletrônicos das revistas tradicionais.

Uma outra subcategoria das revistas eletrônicas é representada pelas *revistas originalmente eletrônicas*, que não nasceram de revistas tradicionalmente baseadas no papel. Exemplos conhecidos desta categoria incluem as revistas JAIR (*Journal of Artificial Intelligence Research*) e MISQ Discovery. Apesar de ambas as revistas empregarem a revisão pelos pares tradicional, JAIR abandonou a organização da revista por edições, e a revista MISQ Discovery introduziu os conceitos de publicações “arquivadas”, que somente podem ser editadas para correções de gramática, digitação ou atualização de links, e “ativas”, que podem ser atualizadas sempre pelos autores.

Uma terceira subcategoria das revistas eletrônicas é representada pelos repositórios eletrônicos ou “*pre-print archives*”, cuja importância na transição para as novas formas de publicação já foi levantada na seção 2.3.2 deste capítulo.

Uma outra categoria de infra-estrutura de conhecimento é representada pelas *bibliotecas digitais*, que objetivam “trazer a biblioteca até o usuário” (STORY et al., 1992), que organizam e oferecem acesso a grandes coleções de documentos. Um exemplo conhecido é a biblioteca digital JSTOR, que em 2002 contava com mais de 1500 instituições de pesquisa participantes e oferecia acesso ao conteúdo de mais de 300 revistas em 15 disciplinas (HARS, 2003).

Uma alternativa que se mostra atrativa às revistas eletrônicas são as *enciclopédias*, infra-estruturas que focam na organização do conhecimento de um assunto ou grupos de assuntos. Duas diferenças importantes entre as revistas e as enciclopédias é que as enciclopédias objetivam apresentar o conhecimento de um domínio de forma organizada e com extensivas referências cruzadas entre suas entradas; a outra diferença está na maneira como a autoria de um nó de conhecimento é tratada. Em uma revista científica, os autores de um artigo são conhecidos, e considera-se como principal papel dos autores a divulgação do conhecimento novo. Já nas enciclopédias, os autores de cada entrada permanecem no anonimato, e assume-se que o principal papel deles é organizar o conhecimento, monitorando um determinado campo constantemente à procura de informações novas ou ausentes. Exemplos de enciclopédias eletrônicas incluem os portais *Free Online Dictionary of Computing*, *Internet Encyclopedia of Philosophy* e a base de dados *GenBank*. Uma descrição destes portais pode ser encontrada, por exemplo, em HARS (2003).

Colaboratórios representam uma outra categoria de infra-estrutura de conhecimento, e este termo é uma combinação das palavras “colaboração” e “laboratório”. Em contraste com as revistas eletrônicas e as enciclopédias, que focam fundamentalmente na disseminação do conhecimento, os colaboratórios suportam diretamente a pesquisa científica, o processo de criação do conhecimento e a comunicação entre os pesquisadores. BARUA et al. (1995) descrevem um colaboratório como “uma plataforma eletrônica e aberta que permite que indivíduos ou grupos com interesses em comum possam criar, trocar e disseminar idéias e conhecimento”. Colaboratórios são altamente interativos e muito mais fluidos no seu conteúdo (HARS, 2003). Exemplos incluem o portal baseado na Web *Principia Cybernetica Web*, o experimental *Information Systems Cybrarium* e o colaboratório de Física Espacial e Aeronáutica da *National Science Foundation* (NSF).

Bulletin boards são “sistemas que permitem que usuários submetam documentos a um quadro branco eletrônico que pode ser visto por outros usuários” (HARS, 2003). *Newsgroups* são um caso especial deles, que permitem a anotação dos documentos na forma de linhas de discussão. Já os *list servers* utilizam a política de distribuir os documentos aos leitores. *Bulletin boards* tornaram-se uma das formas mais difundidas de comunicação eletrônica.

Finalmente, a última categoria que discutiremos refere-se aos *servidores lógicos* ou, em inglês, “*logic servers*”. Nesta categoria incluem-se as infra-estruturas que focam em outros mecanismos de representação do conhecimento que não sejam artigos em texto convencionais. Exemplos destas infra-estruturas incluem o servidor *Ontolingua*, desenvolvido em Stanford, que permite aos usuários registrados criar, através de uma interface baseada na Web, suas próprias ontologias, além de poderem visualizar ontologias de outros usuários. Outros exemplos desta categoria são o portal CYC, no qual o conhecimento é organizado como um “mar de axiomas” (CYCORP, 1999), e o servidor *Toronto Virtual Enterprise* (TOVE), com características semelhantes ao Ontolingua.

2.4.2 Revistas eletrônicas - modelos emergentes de publicação científica

Nesta seção faremos uma análise comparativa de alguns dos modelos emergentes mais significativos propostos nos últimos anos. Será apresentada

inicialmente a proposta mais discutível - conhecida como *non-refereeing* - e em seguida serão apresentados os modelos das revistas *BMJ*, *Psycoloquy*, *BBS*, *PubMed Central*, *JIME*, o modelo da Física de Alta Energia e, finalmente, o modelo hipotético DJ proposto por John Smith.

2.4.2.1 Publicação sem revisão (*non-refereeing*)

Trata-se da mais radical das propostas de mudança no processo de revisão da literatura. Seus proponentes defendem a exclusão completa do *peer-review* do processo de comunicação científica, e o fim das revistas – no ambiente da Internet os autores podem publicar todo e qualquer material, deixando aos leitores a tarefa de filtrá-los. Todavia, existe um consenso na comunidade científica da importância da revisão pelos pares na garantia de qualidade da pesquisa científica publicada e preservada para a posteridade. Além disso, em áreas como Medicina e a Química, por exemplo, esta abordagem se torna inaceitável em certas questões, sobretudo as que envolvem risco de vida. Desta forma, dentre as propostas de mudanças no modelo de publicação científica, esta é sem dúvida a que apresenta menor probabilidade de concretizar-se.

Uma abordagem bem mais suave do que esta, implantada desde o ano de 2001 em algumas áreas da Física, defende a manutenção do *peer-review*, mas a publicação de manuscritos mesmo antes de sua revisão. Esta abordagem é conhecida como **publicação instantânea** (HARS, 2003). Contra essa abordagem existe o argumento de que, com o aumento exponencial do volume de publicações (ODLYZKO, 1997), fica difícil encontrar o bom artigo sem o “filtro” do *peer-review*. Mas este argumento, de acordo com HARS (2003), assume incorretamente que o leitor não é capaz de distinguir entre artigos não revisados, e artigos revisados por uma revista científica de nome.

2.4.2.2 O modelo da revista BMJ

A revista britânica *BMJ* (*British Medical Journal*) está experimentando desde o ano de 1999 o sistema *aberto* de revisão da literatura médica, em que os autores conhecem os nomes dos revisores de seus artigos e dos editores da casa. No passado a revista já utilizou também o sistema fechado de revisão da literatura.

O processo de avaliação funciona da seguinte forma. Uma vez submetido um artigo, ele passa por um filtro inicial realizado geralmente por dois editores médicos da casa. Nesta fase inicial, aproximadamente metade dos artigos submetidos é rejeitada por

uma série de fatores, entre os quais a falta de originalidade, a ausência de uma mensagem significativa para o público médico, a presença de falhas científicas graves etc. Esta etapa em geral não dura mais do que duas semanas.

Os artigos que sobrevivem ao filtro inicial são enviados para um ou mais revisores externos selecionados de um banco de dados formado por mais de 4000 especialistas. Na fase seguinte, os artigos aprovados por esses revisores externos são submetidos a um dos *comitês editoriais* da revista. Um comitê geralmente é constituído de dois clínicos, um ou dois editores e um técnico em estatística. Existem comitês específicos para determinadas classes de artigos. Alguns artigos podem ainda ser submetidos à avaliação de um *comitê de ética*. De acordo com os editores da revista BMJ, a decisão final no que diz respeito à aceitação ou rejeição do artigo geralmente é concluída em até oito semanas da data de submissão.

Quando o artigo é aceito sujeito a revisões, os autores são incumbidos de retorná-los já revisados num prazo de até dois meses. A publicação acontece geralmente em até três meses da data de aceitação.

Como planos futuros, os editores da revista pretendem a partir de um dado momento começar a listar os nomes dos revisores no final dos artigos. Este seria um primeiro passo para mover a revista em direção a um novo sistema, em que autores e leitores acompanhariam passo a passo o processo de revisão através da Web, e participariam ativamente dele através de comentários e anotações. Segundo SMITH (1999a), “o processo de revisão se tornará gradativamente um discurso científico, e não mais um julgamento sumário”.

2.4.2.3 O modelo da revista BBS

A revista *Behavioral and Brain Sciences* (BBS) foi criada por Steven Harnad, pesquisador da universidade de Princeton nos EUA. De acordo com HARNAD (1996), BBS é uma revista no formato impresso especializada em comentários abertos, que ele chama de "*open peer-commentary*". O processo de revisão é realizado tanto antes da publicação, com o *peer-review* tradicional, quanto após a publicação, com o *peer-commentary*.

Se um artigo passa pelo filtro do *peer-review* e é aceito, ele passa a circular entre especialistas distribuídos ao redor de todo o mundo. Existem 100 comentaristas potenciais, que são convidados a comentar os artigos (HARNAD, 1996). O autor do

artigo deve responder aos comentários, criando linhas de discussão. Em seguida, o artigo é publicado em conjunto com 20 a 30 comentários revisados, mais as respostas do autor aos comentários.

O modelo da BBS tem tido grande aceitação na comunidade das ciências cognitivas e comportamentais: no seu 6º ano de funcionamento, a revista já apresentava o terceiro maior fator de impacto entre 1200 revistas indexadas na *Social Science Citation Index* (HARNAD, 1996). Todavia, a revista apresenta algumas limitações, como o próprio Harnad enaltece. O tempo é um deles. Como se era de esperar, o *peer-review* é um fator de retardo, uma vez que é feito da maneira tradicional antes da publicação. Mesmo que o cronômetro fosse zerado no momento em que o artigo é aceito, este seria publicado no formato impresso dentro do período de 4 a 6 semanas.

2.4.2.4 O modelo da revista *Psychology*

Psychology é uma revista totalmente eletrônica na área de Psicologia, criada em 1990 pelo pesquisador Harnad, o mesmo responsável pela revista BBS. No modelo de Harnad, no qual a revisão é realizada de forma aberta (*open peer-review*), um artigo é revisado através de comentários de outros autores e/ou revisores e das respostas deles a estes comentários.

Todos os comentários feitos em cima de um determinado artigo são revisados por um membro da revista, que possui uma equipe de 35 revisores. A revista implementa uma forma inovadora de comunicação, em que os autores expõem suas idéias e solicitam *feedback* por parte de especialistas não apenas na área da psicologia, como em disciplinas relacionadas. A revisão de cada item exposto e de cada item de resposta é realizada de forma extremamente rápida, algumas vezes dentro de poucas horas. A esta forma de comunicação científica Harnad convencionou chamar de “*scholarly skywriting*”, em alusão ao fato de que tudo acontece como se cada contribuição estivesse sendo escrita no céu, estando sempre visíveis aos participantes do processo.

De acordo com HARNAD (1990), “*skywriting* promete trazer a velocidade da comunicação científica muito próxima da velocidade do pensamento (...), conduzida pela disciplina da mídia escrita, monitorada pelo *peer-review*, e arquivada permanentemente para referências futuras”.

A revista aparece em duas versões distintas, uma direcionada à grande rede através de Princeton, e a outra armazenada na universidade de Tulane e arquivada na universidade de Houston. Qualquer pessoa pode se inscrever gratuitamente na revista e se tornar um leitor, bastando para isso enviar uma mensagem eletrônica para um dos endereços disponíveis.

Um fato curioso sobre a revista apontado por HARNAD (1990) é que praticamente todo o trabalho realizado em prol da revista tem sido feito de graça pelos envolvidos. A equipe de especialistas responsável pela revisão não recebe nenhuma gratificação pelo serviço prestado; as universidades envolvidas oferecem os recursos computacionais gratuitamente; o modesto subsídio da Associação Americana de Psicologia (APA) é utilizado unicamente para pagar uma pessoa para manter a lista de endereços eletrônicos dos membros e publicar os comentários e contribuições na Web. *Psychology* foi selecionada em uma pesquisa realizada anualmente pela revista *Library Journal* como uma das melhores revistas emergentes em 1990.

2.4.2.5 O modelo da revista PubMed Central

A revista PMC (*PubMed Central*) é publicada na Web desde o ano de 2000 pelo NIH (*National Institutes of Health*). A proposta da revista é de funcionar como um repositório eletrônico e aberto de artigos revisados por revistas selecionadas na área de ciências naturais, promovendo o intercâmbio de idéias entre autores e leitores (KLING et al., 2001).

Desde o dia em que entrou no ar, a revista PMC somente aceita relatórios técnicos ou pré-publicações que tenham sido submetidos ao processo de *peer-review*. Além da exigência de que o artigo a ser publicado tenha sido revisado por alguma revista selecionada, PMC também adota um processo de revisão independente necessário para compatibilizar o artigo com os padrões de qualidade da revista.

Quando uma revista submete um artigo à PMC, o processo se inicia através de uma revisão automática e também manual do material submetido, com o objetivo de garantir que o documento tenha a mesma acurácia da versão original, publicada no sítio da revista que está submetendo o artigo. Os artigos submetidos devem estar de acordo com os padrões de completude e corretude sintática, estabelecidos pela revista PMC para garantir a integridade do repositório. Todo artigo deve ser enviado no formato XML utilizando um DTD estabelecido pela PMC, com o objetivo de normalizar a

marcação das diversas partes do documento, como os autores, afiliações, seções do artigo, referências etc. Se erros são encontrados durante essa fase de revisão, a revista que está submetendo o artigo é solicitada a efetuar as devidas correções.

Muitas revistas não liberam o texto completo de seus artigos para repositórios como a PMC, ou liberam com um atraso de meses ou até 1 ano. Esta prática se tornou bastante comum nos últimos anos, e vem sendo adotada e defendida pela associação americana de bibliotecas acadêmicas (AAAS). HARNAD (2001b) comenta que “disponibilizar voluntariamente o conteúdo (das revistas) com um atraso de 12 meses é certamente um passo bem-vindo por parte da AAAS, mas isto é muito pouco e muito tarde...”.

A revista recentemente relaxou a exigência de que o texto completo fosse disponibilizado no sítio da PMC. Caso uma revista queira publicar o texto completo de seus artigos exclusivamente em seu sítio particular, a PMC colocará um apontador para o sítio da revista. De qualquer maneira, estando ou não o texto completo disponível no sítio da PMC, esta última indexa todos os documentos e disponibiliza um serviço de busca aos leitores da revista. A revista PubMed Central, no entanto, se dá o direito de exibir o texto completo de artigos que não aparecem gratuitamente até no máximo 1 ano na página da revista. Em outras palavras, se uma revista demora mais de um ano para disponibilizar o texto completo de um artigo no seu sítio ou no sítio da PMC, esta se dá o direito de exibir o texto do artigo em seu repositório de forma gratuita aos leitores.

2.4.2.6 O modelo da revista *JIME*

Publicada na Web desde o ano de 1996, *JIME (Journal of Interactive Media in Education)* é uma revista totalmente eletrônica e gratuita, criada com o objetivo de incentivar a pesquisa na área de multimídia aplicada à educação. No modelo de publicação do *JIME*, após ser submetido, um artigo passa por duas fases de revisão *aberta* pelos pares: a primeira delas *privada*, em que os revisores comentam e os autores respondem aos comentários, e a segunda *pública*, onde os leitores também contribuem.

Algumas características inovadoras merecem destaque neste modelo. Primeiro, os atores do processo de publicação têm seus papéis modificados: autores respondem aos comentários dos revisores, leitores podem contribuir na discussão pública etc. Segundo, o debate em cima do artigo se inicia tão logo ele entre em revisão, continuando mesmo após sua publicação, enquanto na grande maioria das revistas atuais

o debate é iniciado apenas quando de sua publicação. No JIME, o debate é integrado ao documento, através do ambiente D3E, em contraste com outras revistas em que o debate fica fragmentado em páginas pessoais, em listas de discussão etc. O ciclo de vida de um artigo submetido ao JIME é mostrado na figura a seguir:

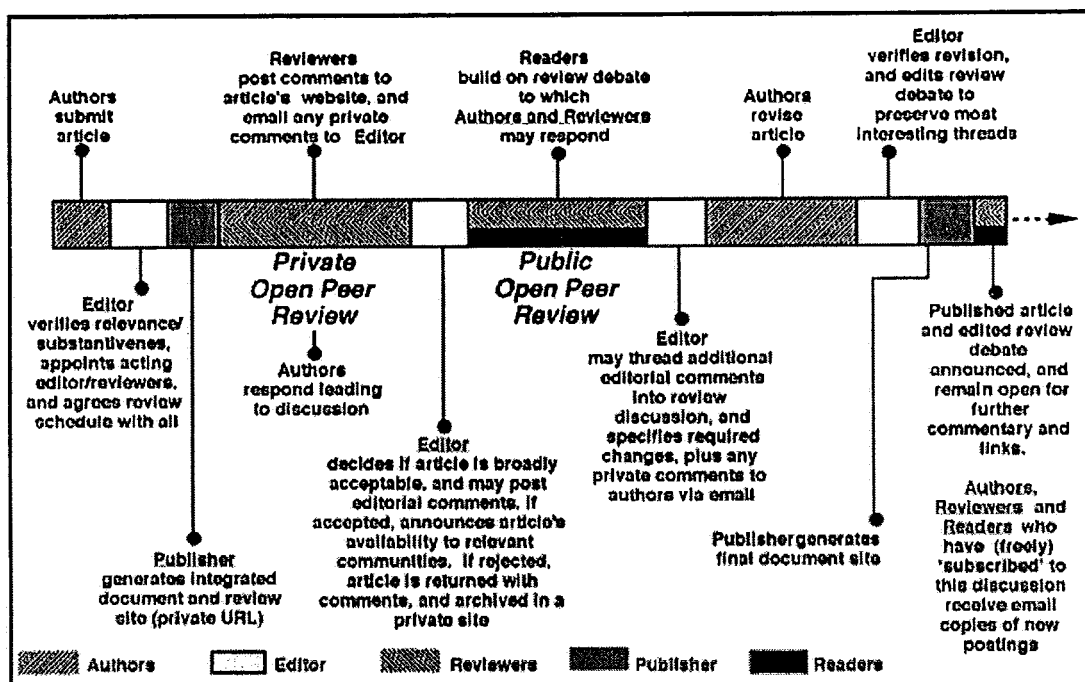


Figura 2.4 – Ciclo de vida de um artigo submetido ao JIME

Terceiro, a revista permite, através de sua interface, que seus documentos sejam “incrementados” pelos usuários (leitores, revisores, autores), que podem adicionar demonstrações interativas, clipes de áudio e vídeo, discussões e apontadores que ligam o documento a outros trabalhos relacionados. Todavia, documentos são submetidos e armazenados no ambiente D3E sempre no padrão HTML, uma formato limitado já que mistura o conteúdo do artigo com a apresentação. Sendo assim, não é possível customizar a visualização de um documento para contextos e públicos diferentes. Finalmente, os comentários feitos pelos leitores e revisores são sempre em cima do documento como um todo: não é possível comentar uma seção, um parágrafo ou um trecho específico do documento.

O trabalho de pesquisa em cima do JIME é exploratório, já que não existe uma teoria consistente por trás do modelo proposto. De acordo com os criadores do JIME, a revista não foi criada com a preocupação de atender a todas as disciplinas do conhecimento, mas àquelas em que os critérios de aceitação/rejeição devem ser

negociados (SHUM & SUMNER, 2001). Eles reconhecem as limitações de aplicar a metodologia do JIME, sobretudo no que se refere ao discurso científico, a determinados campos do conhecimento em que não existe a cultura da argumentação, do discurso aberto e da negociação. A figura a seguir ilustra um artigo publicado no JIME. Observe que o debate em cima do artigo é exibido no frame à direita:

The screenshot displays the JIME interface. On the left is a navigation pane with a table of contents. The main content area shows the article '2. The Media Cube' by Mark G. Draping, Susan A. Gabe, and others. The article text includes a paragraph about the Media Cube framework and a 3D diagram. The diagram is a cube with three axes: 'Richness' (vertical), 'Interact' (horizontal), and 'Accessibility' (depth). Points on the axes are labeled with numbers 1 through 8. Point 1 is at the origin, point 2 is on the Richness axis, point 3 is on the Interact axis, point 4 is on the Accessibility axis, point 5 is at the end of the Richness axis, point 6 is at the end of the Interact axis, point 7 is at the end of the Accessibility axis, and point 8 is at the top-right-back corner. The article text includes a paragraph: 'Even though learning is complex in these media dimensions, we do not explicitly discuss the effects have on knowledge construction, but for the nearly effects these dimensions have an influence on construction.' Below the diagram is a caption: 'Figure 4: The Media Cube'. The right sidebar shows a list of reviews, with the first one titled 'Re: 2. The Media Cube' and the second 'Re: 2. The Media Cube framework'. The first review includes the text: 'I have some concerns about the rendering of specific technologies as points rather than as spaces within the cube. In particular, the characterization of immersive VR of the "passive absorption" point on the size activity dimension is misleading, certainly that is true for simple such applications (such as educational walk-through), but there are now several examples of applications which could reasonably be classified as other active applications or even constructions.'

Figura 2.5 – Exemplo de documento sob revisão no JIME

2.4.2.7 O modelo da Física de Alta Energia

Neste modelo, o autor carrega seu artigo em um repositório de *e-print* e solicita comentários de outros físicos. Com base nos comentários, o autor revisa e aprimora o artigo. O artigo é então avaliado na maneira convencional (*peer-review*) através de alguma revista. Finalmente, a versão antiga do artigo dá lugar à sua nova versão, que passa a ser identificada como *publicação*. Um ponto importante é que todo o material que aparece como pré-publicação nos repositórios é não-revisado, já que a revisão formal é realizada em uma fase seguinte. No entanto, os artigos ganham uma estampa

de *pré-publicação* para alertar os leitores do fato de que o documento ainda não foi submetido ao processo de validação.

É exatamente isto que está acontecendo na comunidade da *Física de Alta Energia*, desde o ano de 1991, quando o físico Paul Ginsparg introduziu o sistema. Ele é o principal responsável pela criação do repositório de e-print arxiv.org.

Em HARNAD (2001a), o autor comenta que "os mesmos autores que disponibilizam gratuitamente seus artigos na Web continuam a submetê-los às revistas de sua escolha, da mesma maneira como faziam antigamente, e praticamente todos os artigos no repositório aparecem nas revistas doze meses após a submissão". Desta maneira, a fase de revisão da literatura, um dos alicerces do modelo corrente de publicação científica, permaneceu intacta na Física, já que continua sendo realizada da maneira tradicional pelas revistas.

Em função desta iniciativa da comunidade física, nos últimos anos ganhou força um movimento em defesa dos repositórios abertos, conhecido como *Open Archives Initiative* (OAI). O principal objetivo é tornar o conteúdo dos repositórios de literatura acadêmica interoperáveis, criando um repositório único virtual. Uma das ações do movimento é oferecer software livre e gratuito para instituições que desejem criar seus repositórios seguindo o padrão OAI. Para atingir esses objetivos, foram criados padrões de metadados de forma a tornar o conteúdo dessas instituições interligado, possibilitando ainda uma busca integrada ao repositório virtual global (HARNAD, 2001a).

Ainda de acordo com HARNAD (2001a), a criação desses repositórios distribuídos e interoperáveis beneficia as instituições de pesquisa de três maneiras. Primeiro, maximiza a visibilidade e o impacto da pesquisa revisada da instituição. Segundo, maximiza o acesso pelos pesquisadores ao resultado da pesquisa revisada de todas as outras instituições. Terceiro, as instituições de pesquisa podem deflagrar a transição do modelo antigo de publicação ao modelo OAI, reduzindo sensivelmente os gastos anuais com assinaturas de revistas.

Esta abordagem, utilizada com sucesso como modelo para disciplinas próximas da Física, como Matemática e Ciência da Computação, tem sido vista com ressalvas em outros campos. De acordo com ROWLAND (2002), "existem diferenças significativas de atitude entre as culturas de diferentes disciplinas". Um dos fatores de sucesso do modelo na comunidade física é a tradição de compartilhamento de manuscritos de artigos entre os pesquisadores.

2.4.2.8 O modelo hipotético DJ

Em SMITH (1999b), John Smith descreve o modelo de funcionamento de uma revista hipotética que ele chama de *Deconstructed Journal* (DJ). Como o próprio autor afirma, este modelo se baseia em grande parte nas idéias de Paul Ginsparg, principal responsável pelo modelo da Física de Alta Energia e criador dos repositórios de *e-print* da mesma comunidade.

De forma resumida, a revista funcionaria da seguinte forma. O autor escreve um artigo, carrega-o em algum repositório online e notifica uma ou mais *Organizações Avaliadoras* (OAs) da sua existência. Uma ou mais OAs revisam o artigo, solicitam alterações da forma convencional e eventualmente aprovam o artigo. O autor revisa o artigo a partir dos comentários feitos pelas OAs. Em seguida o autor notifica elementos aos quais Smith convencionou chamar de "*Subject Focal Points*" (SFPs). Os SFPs selecionam os artigos mais relevantes às suas áreas e criam apontadores para os artigos incluídos.

No modelo de Smith, um dos elementos centrais proposto é de fato o *Subject Focal Point* (SFP). Seu principal papel é atuar como um "filtro" entre o conteúdo da Internet e o usuário assinante, e não funcionar como um repositório da literatura. Os coordenadores do SFP em geral não atuam no processo de revisão da literatura, papel este que cabe às Organizações Avaliadoras. Smith sugere que as OAs poderiam ser sociedades científicas, instituições de pesquisa ou mesmo universidades. Elas utilizariam revisores pagos ou convidados para a atividade de revisão da literatura. O autor aponta a possibilidade de corrupções nessas organizações, mas considera que as organizações corruptas tenderiam naturalmente a desaparecer. No que se refere ao controle da *forma* do artigo, este seria realizado pelo próprio autor através de ferramentas de apoio à publicação, ou através de especialistas nas universidades.

Uma das vantagens do modelo bastante defendida diz respeito ao *reconhecimento dos autores*, uma função que seria realizada de forma conjunta pelos SFPs e pelas OAs. O modelo permite que um autor tenha seu trabalho avaliado por mais de uma Organização Avaliadora e também que mais de um SFP aponte para o mesmo item, já que os SFPs não detêm direito sobre nenhum item. Isto em princípio minimizaria o problema do atraso causado pelo processo de avaliação, uma vez que mais de uma organização estaria avaliando o mesmo item. Todavia, existe o problema

de um artigo ser aprovado por um SFP e rejeitado por outro. Smith faz ainda algumas ressalvas com relação ao problema da *autenticação* das publicações no modelo DJ. Ele defende que um selo de aprovação teria de ser de alguma forma embutido no artigo, por parte de uma Organização Avaliadora, para garantir aos leitores a qualidade da publicação.

SMITH (1999b) não apresenta informações detalhadas de como funcionaria o processo de revisão propriamente dito, nem mesmo se este processo seria realizado de forma aberta ou fechada. Um dos pontos de destaque deste modelo é a revisão feita pelas OAs de forma independente, o que poderia trazer ganhos consideráveis na velocidade do processo de publicação.

2.5 Características das infra-estruturas de conhecimento científico baseadas na Web

Para analisar as perspectivas futuras no que diz respeito às infra-estruturas do conhecimento, HARS (2003) propõe um framework com cinco grandes grupos de características: 1) Modo de interação; 2) Estruturação do conhecimento; 3) Apresentação; 4) Gerenciamento; e 5) Tecnologia.

Nesta seção iremos descrever em linhas gerais estas características e de que forma elas trariam melhorias significativas para o processo de publicação científica. Desta análise surgiram algumas das principais propostas do ambiente CEJ, foco principal desta monografia, e que serão descritas no capítulo seguinte. Uma tabela ilustrando as características deste *framework*, adaptada de HARS (2003), é apresentada no apêndice A desta monografia.

2.5.1 Modo de interação

O modo de interação descreve de que maneira os atores participantes de uma infra-estrutura do conhecimento com ela interagem. Quanto maior o grau de interação, maior o potencial de obtenção de ciclos menores para a publicação de um artigo. Atualmente, somente uma pequena parcela das infra-estruturas de conhecimento tem procurado explorar as oportunidades de níveis maiores de interatividade oferecidas pela tecnologia da informação, incluindo-se aí o suporte a anotações e discussões baseadas nas contribuições de leitores e revisores. Muitas revistas já oferecem suporte ao processo de revisão, oferecendo geralmente um espaço em separado para o trabalho dos

revisores. Algumas infra-estruturas chegam mesmo a permitir o gerenciamento do *workflow* de revisão. Por último, a maioria das revistas eletrônicas não oferece nenhum tipo de suporte à autoria do artigo através de seus sistemas, exigindo a submissão do artigo já pronto. Para suportar tal funcionalidade, mecanismos de edição online são necessários. Uma das possibilidades a ser explorada seria investigar a API de alguns navegadores Web conhecidos, como, por exemplo, a API do Microsoft Internet Explorer, que atualmente já permite a abertura de um documento HTML em modo de edição.

Para a submissão dos manuscritos, a abordagem predominante e preferida pelas revistas eletrônicas tem sido o correio eletrônico. Neste caso, geralmente o processamento das submissões é realizado manualmente. Algumas revistas permitem também a submissão através do protocolo FTP e outras através do próprio navegador Web, via HTTP. Um número muito reduzido delas já permite a criação e a edição das submissões pelo próprio sistema.

Uma discussão importante diz respeito ao momento de publicação de um manuscrito. A maioria das revistas eletrônicas manteve a abordagem tradicional de só publicar o artigo depois de revisado e no momento de publicação de uma edição (*issue*) da revista. Um número crescente de revistas, no entanto, já publicam seus artigos tão logo eles passem a etapa de revisão. De acordo com HARS (2003), algumas revistas decidiram pela reengenharia do processo de revisão, e começaram a paralelizar as atividades de revisão e publicação. Na ISWorldNet, por exemplo, artigos submetidos são publicados imediatamente, mas em uma seção a parte. Leitores e revisores são convidados a comentar e, caso uma decisão negativa com relação a um artigo seja tomada, o artigo é simplesmente removido da revista.

No ambiente eletrônico, um artigo ou contribuição de um autor em princípio poderia ser atualizado, com ou sem revisão. Na prática, porém, muitas revistas eletrônicas não permitem aos autores editarem suas contribuições, outro resquício da era de Gutenberg – do papel.

2.5.1.1 CSCW e *Groupware*

O apoio ao trabalho cooperativo pode ser de fundamental importância na execução de determinadas fases ou tarefas de um processo de publicação científica, sobretudo na revisão e comentários pelos pares. A colaboração pode apoiar grupos de

trabalho na realização de tarefas comuns, como a revisão de um artigo ou a aprovação do mesmo.

MORAIS (2003) argumenta que o principal objetivo de uma aplicação colaborativa está em oferecer compartilhamento, que pode se expressar em diversos níveis, desde o tecnológico ao social, de acordo com as interações a que se propõem a apoiar. O compartilhamento do ponto de vista tecnológico implica em conectividade, distribuição de conteúdo de trabalho e administração de acesso. O compartilhamento do ponto de vista social implica em comunicação com significado, entendimento e, nos casos em que há um compromisso de interação entre as partes, o encontro ou convergência de pensamentos (POLTROCK & GRUDIN, 1997). Para um ambiente de execução de processos de publicação, é necessário um compartilhamento tanto tecnológico quanto social.

Nesta discussão conjunta dos conceitos de *CSCW* e *groupware*, é importante destacar que estes dois termos frequentemente têm sido utilizados como sinônimos de forma incorreta. Enquanto *CSCW* refere-se à pesquisa na área do trabalho em grupo e à maneira como os computadores podem apoiá-lo, o termo *groupware* tem sido empregado para designar a tecnologia gerada pela pesquisa em *CSCW* (ELLIS, GIBBS et al., 1991; GRUDIN, 1994; QUATERMAN, 1990).

Para CHAFFEY (1998), o trabalho cooperativo baseia-se três funções principais: Comunicação, Colaboração e Cooperação. Comunicação é a função que permite o compartilhamento de informações pelo grupo. Colaboração é a ação de cooperação na solução de problemas empresariais. E coordenação é o controle da efetividade e do cumprimento de objetivos do trabalho por parte da equipe.

CASTRO (2002) identifica como principais requisitos de um sistema baseado em *groupware* os itens a seguir:

- O sistema deve facilitar a cooperação entre indivíduos, aoinvés de impor práticas que causem mudanças radicais na forma de trabalho;
- Sistemas de *CSCW* devem reconhecer que mudanças são frequentes neste contexto, e que, por isso, devem ser capazes de permitir a redefinição de procedimentos e processos, além de disseminar estas mudanças entre os participantes;

- A construção de aplicações menores e inter-relacionadas é preferível ao desenvolvimento de aplicações monolíticas que incluem o conjunto completo de tarefas;
- Informações que serão usadas no trabalho cooperativo precisam estar fora do domínio de um indivíduo.

ELLIS, GIBBS et al. (1991) dividem os sistemas baseados em *groupware* em quatro quadrantes no que se refere às noções de tempo e espaço sob as quais as interações são realizadas, conforme é ilustrado na Figura 2.6:

	Mesmo Momento (síncrono)	Momentos Distintos (assíncrono)
Mesmo local	Salas de Reunião Compartilhamento de Tela/Documentos	Workflow Compartilhamento de documentos
Locais distintos	Compartilhamento de Tela/Documentos Video / Audio Conferência Conferência por computador	WorkFlow Conferência assíncrona pelo computador

Figura 2.6 – Classificação de Groupware (ELLIS, GIBBS et al., 1991)

Quando o envio e o recebimento de mensagens ocorrem simultaneamente, a comunicação é dita síncrona. A comunicação assíncrona ocorre quando o receptor recebe a informação em um momento diferente do momento de envio. Já os fatores tempo e distância avaliam se o sistema permite que os usuários estejam em lugares distintos para estabelecer a comunicação. Interações face-a-face são exemplos de cooperação realizada no mesmo ambiente físico e no mesmo espaço de tempo. Já uma troca de mensagens por correio eletrônico se dá em diferentes períodos de tempo e com as partes estando em diferentes ambientes.

2.5.1.2 Mecanismos de *Awareness*

A área de CSCW vem buscando, desde os seus primórdios, um meio de suportar adequadamente o trabalho em equipe. Um dos principais problemas que tem se apresentado é a falta de contexto entre os participantes, que ocorre quando os membros de um grupo de trabalho desconhecem o que seus colegas estão fazendo, ou não sabem onde suas atividades se encaixam no trabalho como um todo, nem qual é a situação desse trabalho. Esse contexto é um ponto importante em sistemas cooperativos, estendendo-se para não somente o conteúdo das contribuições individuais, mas também ao seu significado para o grupo como um todo e seu objetivo (PEREIRA et al, 2004). O fornecimento deste contexto aos membros de um grupo é chamado *awareness* ou percepção.

A definição de *awareness* para trabalho cooperativo é “o entendimento das atividades dos outros, na qual provê um contexto para a sua própria atividade” (GUTWIN & GREENBERG, 1998). Existe a necessidade de prover mecanismos de *awareness* tanto para aplicações síncronas quanto assíncronas, pois nos dois contextos percebe-se que uma pessoa normalmente não consegue fazer contribuições valiosas para um grupo, se a mesma não conhece informações prévias que são concentradas no assunto correspondente (MORAIS, 2003).

Em um ambiente de publicação científica, mecanismos de percepção são fundamentais para garantir que os participantes de um processo de publicação percebam e compreendam no que os demais membros estão trabalhando.

Desta forma, mecanismos de notificação representam um recurso valioso de interação assíncrona entre os atores e as infra-estruturas de conhecimento científico. Neste contexto, podemos separar as revistas eletrônicas em duas categorias: passivas e ativas. No primeiro caso, as revistas esperam que o leitor tome iniciativa na recuperação da informação disponível no portal da revista, enquanto no segundo caso a revista notifica os leitores quando informações relevantes se tornam disponíveis. Por exemplo, quando da submissão de novos documentos à revista, leitores poderiam ser avisados através do correio eletrônico.

Existem diversas alternativas para a implementação de notificações: listas de usuários em que uma mesma mensagem é enviada a todos os inscritos na lista, notificações personalizadas, onde os atores do processo (autores, leitores etc) podem configurar sobre que eventos e sobre quais contribuições desejam ser notificados, e qual a frequência das notificações.

2.5.1.3 Sistemas de *workflows*

O gerenciamento e a execução de processos de publicação científica podem ser conseguidos com o apoio de sistemas de *workflow*. Do ponto de vista tecnológico, *workflow* tem sido visto como um tipo relativamente simples de *groupware*, pois suporta a coordenação de pessoas e processos automatizados, isto é, os processos de negócio (MORAIS, 2003). Um processo de negócio pode ser representado através de uma definição de processo. Segundo a WfMC (ALLEN, 2001), esta definição consiste de uma rede de atividades e seus relacionamentos, critérios de início e término do processo, e informações sobre as atividades em si, como participantes, aplicativos e dados de sistemas de informação relacionados, entre outros. Processos de negócios podem ser representados por fluxos de trabalho, que são modelos que especificam as atividades do processo de negócio, a ordem e as condições em que estas devem ser executadas, os executores de cada atividade, as ferramentas a serem utilizadas e os documentos manipulados durante a sua execução.

Um sistema de gerenciamento de fluxo de trabalho (ou sistema de *workflow*) tem como principais objetivos a automação e o monitoramento de processos representados por fluxos de trabalho. Constitui-se de um conjunto de ferramentas que permite o projeto e a definição de fluxos de trabalho, sua instanciação e execução controlada, bem como a coordenação e a integração de diversas ferramentas dentro de um mesmo fluxo de trabalho (ARAUJO & BORGES, 2001).

Os principais elementos de um fluxo de trabalho são as atividades, que correspondem a uma parcela de trabalho a ser feita dentro do processo. São definidas por elementos como nome, objetivos, instruções, dados, formulários e documentos necessários que podem ser manipulados durante sua execução. As atividades são executadas por indivíduos, agentes, grupos ou papéis. Os indivíduos podem ser responsáveis pela execução de determinada atividade dentro de um fluxo de trabalho. Um papel reúne um conjunto de características ou responsabilidades necessárias para a execução de uma ou mais atividades e, no momento da execução de determinada atividade, são representados por atores que podem ser pessoas ou agentes automatizados.

Um aspecto importante na definição do fluxo de trabalho é o encadeamento de atividades. As atividades em um fluxo de trabalho podem ser executadas de diversas maneiras. KOBIELUS (1997) define quatro tipos de processos, de acordo com o

sequenciamento das atividades que compõem o processo: sequenciais, condicionais, paralelos e concorrentes. Nos processos sequenciais, as atividades são executadas uma após a outra, enquanto nos condicionais existem pontos de decisão, onde um evento é avaliado e isto definirá qual será a próxima atividade a ser iniciada. Mas nos processos condicionais, somente uma atividade poderá ser iniciada como consequência do evento. No caso contrário, o processo é definido como paralelo, no qual diversas atividades podem ser executadas ao mesmo tempo. Processos concorrentes são semelhantes aos paralelos, mas neste caso as atividades paralelas deverão ser iniciadas simultaneamente e concluídas da mesma forma. A figura 2.7, extraída de KOBIELUS (1997), ilustra estes quatro tipos de processos:

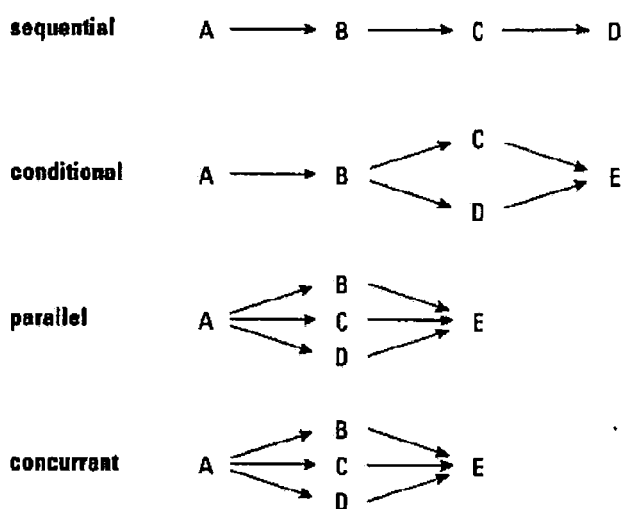


Figura 2.7 – Quatro tipos de processos (KOBIELUS, 1997)

Para tanto, são definidas as rotas, que determinam o sequenciamento de atividades e elementos especiais que indicam condições para que uma atividade seja iniciada.

No modelo tradicional de publicação, conforme mencionado na proposição #10 apresentada na seção 2.2.2.3, as principais atividades do processo de publicação são executadas sequencialmente: a revisão deve ocorrer antes da impressão, que deve acontecer antes da distribuição, que deve ocorrer antes da leitura. No ambiente eletrônico, por outro lado, a impressão é eliminada e a distribuição se torna um aspecto da leitura. Mas vimos que no ambiente eletrônico o espaço é abundante e os artigos podem ser modificados depois de publicados, além de ser tecnicamente possível remover um artigo (o apontador, na verdade) de uma revista, mesmo depois de

publicado. Desta forma, seria plenamente possível publicar um artigo primeiro e depois revisá-lo, de forma que a revisão e a leitura do artigo seriam executadas em paralelo.

Além de prover uma forma de controlar os processos de publicação, um sistema de *workflow* pode ter outras funcionalidades importantes de acompanhamento dos fluxos, como a visualização do status de execução de uma instância ou caso de processo.

KOULOPOULOS (1995) classifica os sistemas de *workflows* de acordo com os métodos de desenvolvimento e os modelos de processos. Os métodos de desenvolvimento são quatro: *ad-hoc*, baseados em transações, baseados em objetos, e baseados em conhecimento. Os modelos de processo são três: baseados em email, baseados em documentos e baseados em processos. Estas categorias se sobrepõem e não devem ser consideradas como estanques (MORAIS, 2003), e é a combinação delas que determina cada um dos tipos de aplicações atendidas pelos sistemas de workflow (VALADARES, 2001).

2.5.2 Estruturação do conhecimento

Organizar o conhecimento e garantir sua consistência ao longo do tempo constitui a segunda dimensão da caracterização das infra-estruturas de conhecimento proposta por HARS (2003).

Muitas revistas eletrônicas ainda adotam a tradicional estruturação dos documentos publicados por volumes e edições (*issues*), muito embora os artigos publicados num mesmo volume ou edição não compartilhem nenhuma característica semelhante a não ser a data de publicação! Muitas revistas emergentes nos últimos anos, todavia, passaram a organizar as publicações em categorias, oferecendo aos leitores uma interface de navegação que possibilita a fácil identificação de artigos pertencentes a um mesmo assunto ou linha de pesquisa.

Uma característica importante diz respeito ao tipo de conteúdo suportado pela infra-estrutura de conhecimento. Revistas tradicionais que se tornaram também eletrônicas costumam separar páginas com o texto completo dos artigos, geralmente acessíveis somente a usuários pagantes, e páginas com somente o resumo (*abstract*) do artigo, normalmente acessíveis ao público em geral. Algumas poucas revistas oferecem acesso a páginas que apresentam outros elementos especiais da pesquisa científica, como dados, eventos e descrições de problemas.

Um outro aspecto relevante diz respeito ao direcionamento dos apontadores. Tradicionalmente, revistas baseadas no papel utilizam apontadores unidirecionais, já que uma citação sempre identifica um outro documento, mas dado um documento, não conseguimos identificar que documentos citaram o mesmo. A Web também trabalha com o conceito de apontadores unidirecionais. Uma possibilidade a ser explorada pelas infra-estruturas emergentes é a de apontadores bidirecionais, que possibilitaria a travessia de um apontador em ambas as direções.

Questões referentes à mudança e manutenção do conhecimento também precisam ser consideradas. Quanto um artigo novo é inserido na base, ou um artigo já na base é atualizado, a mudança duplica ou contradiz o conhecimento já existente? No ambiente eletrônico, checagens automáticas ou semi-automáticas podem ser implementadas com o objetivo de detectar possíveis inconsistências.

Infra-estruturas de conhecimento podem ser classificadas em duas classes: fechadas ou abertas. No primeiro caso, assume-se que todos os artigos estão fisicamente armazenados nos servidores da infra-estrutura. No segundo caso, assume-se que a localização do artigo pode ser qualquer uma, e a inclusão do artigo se dá através do armazenamento da URL que hospeda o mesmo. Infra-estruturas abertas possibilitam a criação de bases de conhecimento maiores e mais abrangentes, mas apresentam como problema a incerteza com relação à disponibilidade do artigo. Infra-estruturas fechadas, por outro lado, possuem maior controle sobre os artigos.

A última característica a ser destacada dentro da dimensão *estruturação do conhecimento* é o versionamento das contribuições. Enquanto nas revistas em papel o artigo é estático, nas revistas eletrônicas pode-se pensar num mecanismo de versionamento. A abordagem mais simples para manter uma publicação ativa consiste em substituir a versão velha do artigo pela nova, e é usada, por exemplo, pela Enciclopédia de Filosofia de Stanford. HARS (2003) apresenta ainda duas outras abordagens: publicações “arquivadas”, que não são mais atualizadas, e publicações “aposentadas”, que também permanecem disponíveis mas não são válidas ou relevantes.

2.5.3 Apresentação

Mecanismos de apresentação podem focar em dois aspectos: apresentação dos componentes internos de um artigo, e apresentação das associações do artigo com outros objetos de conhecimento em que o artigo se encaixa (HARS, 2003).

A primeira característica que discutiremos refere-se às funcionalidades de apoio à visualização do artigo (em inglês, “visualization aids”). Muitas revistas oferecem mapas do sítio que exibem graficamente a estrutura de um artigo. Uma discussão mais aprofundada sobre esse assunto será feita na seção 2.4.2.5, que discute as tecnologias de busca.

Muitas revistas também oferecem funcionalidades de apoio à navegação (“navigation aids”). Exemplos dessas funcionalidades incluem listas com os artigos mais freqüentemente acessados; páginas que exibem os artigos de acordo com sua popularidade etc. Algumas infra-estruturas oferecem a navegação aleatória (*random node functionality*), que permitem a usuários novatos explorar sua base de conhecimento mesmo quando eles desconhecem seu conteúdo.

Mecanismos de filtragem também são importantes na exibição das contribuições em uma infra-estrutura de conhecimento, podendo ser pré-definidos ou customizados pelo usuário. Podem basear-se nas características das contribuições como contagem de acessos, *ratings*, tipos de artigos etc.

Um outro aspecto importante da dimensão “apresentação” é o suporte à **internacionalização**, isto é, à exibição das páginas do sistema e dos nós de conhecimento em diferentes línguas. De acordo com HARS (2003), a maioria das infra-estruturas suporta somente uma língua. Para garantir o requisito de internacionalização, os sistemas deveriam permitir a exibição do conteúdo, sua busca e navegação em uma das línguas familiares ao leitor ou usuário da infra-estrutura.

Considerando o universo de 24000 revistas eletrônicas (HARNAD, 2004), na grande maioria delas os artigos científicos são submetidos nos formatos PDF (*Portable Document Format*), PS (*PostScript*) ou HTML (*Hipertext Markup Language*). O formato HTML tem a vantagem de não necessitar de nenhum software além de um navegador Web para a exibição de um documento. Muitos processadores de texto suportam a tradução dos documentos para o formato HTML. Além disso, a linguagem oferece suporte à formatação de diversos elementos de um documento, como tamanho e cor da fonte, tamanho da margem esquerda etc. Por outro lado, o formato mais empregado é sem dúvida alguma o PDF, que necessita do software proprietário da Adobe para a edição dos documentos, mas oferece a vantagem de preservar o *layout* original do documento, por ser orientado a páginas. Documentos em PDF podem ser acessados através de um navegador Web, assim como o HTML, mas isto requer a instalação de um *plugin* adicional para que possam ser lidos. Além disso, este formato

não se integra com as principais tecnologias da Web, incluindo a linguagem JavaScript e o modelo de objetos do documento (DOM) proposto pela W3C. Finalmente, o formato PS é similar ao PDF, mas não pode ser visualizado através de um navegador Web.

Nos últimos anos, o formato XML tem se consolidado como o principal padrão para o intercâmbio de documentos eletrônicos através da Internet. Muitos projetos de desenvolvimento de software adotaram o XML como principal formato para a documentação. Este formato semi-estruturado permite guardar a semântica de cada “pedaço” de texto, possibilitando que um artigo seja quebrado em suas partes fundamentais: seu título, seus autores, afiliações, seções, parágrafos, figuras, referências etc. Isto permitiria uma busca textual com maior contexto, onde os leitores poderiam procurar por uma palavra “ontologia” somente no resumo dos artigos, ou na seção de conclusão, por exemplo. Adicionalmente, em contraste com o HTML, o formato XML permite a separação do conteúdo de sua apresentação.

O último item que discutiremos ainda dentro da dimensão “apresentação” constitui-se em uma das mais importantes características a ser considerada na discussão de uma infra-estrutura de conhecimento científica: a **separação do conteúdo da apresentação**. Nas revistas tradicionais, baseadas no papel, a exibição do artigo é definida pelo autor e apresentada aos leitores e outras audiências sempre da mesma maneira. Por outro lado, no ambiente eletrônico, a separação do conteúdo da apresentação torna-se uma realidade, graças aos mecanismos oferecido pela tecnologia da informação, como exibições de um mesmo conteúdo de forma não-linear, dinâmica e sensíveis ao contexto, isto é, a apresentação do artigo também poderá variar de acordo com a fase do processo de publicação. Por exemplo, em uma fase de revisão, alguns elementos do artigo seriam destacados ou descritos em maior detalhe, enquanto que numa navegação casual sobre o documento, alguns elementos seriam exibidos e outros escondidos, de acordo com o perfil do leitor.

2.5.4 Gerenciamento de uma infra-estrutura

Esta dimensão contempla os aspectos organizacionais da infra-estrutura de conhecimento. As principais características pertencentes a esta dimensão são a seletividade, a autenticação, o suporte a contribuições anônimas e o modelo de financiamento da infra-estrutura.

A seletividade refere-se à audiência que pode acessar e participar de alguma forma dos serviços oferecidos pela infra-estrutura. Algumas infra-estruturas oferecem acesso ao público em geral, enquanto outras restringem o acesso a usuários qualificados – tipicamente pesquisadores – apesar de a maioria delas oferecer inscrições a usuários comuns, condicionadas ao pagamento de uma taxa. É o caso de sociedades profissionais como a ACM e o IEEE.

A **autenticação** dos usuários também diferencia as revistas eletrônicas e demais infra-estruturas de conhecimento. Algumas não exigem nenhum tipo de autenticação, geralmente as que também não cobram pelo serviço. Todavia, a ausência de autenticação impossibilita a criação de perfis para os usuários que acessam o sistema. A abordagem mais comum é a infra-estrutura oferecer algumas funcionalidades sem autenticação, e outras funcionalidades exigindo autenticação, por exemplo, para as atividades de submissão e revisão de artigos.

Para implementar a autenticação dos usuários, algumas alternativas podem ser consideradas. Uma alternativa comum é através da criação de um nome de usuário (*login*) e uma senha para cada usuário, exigindo-se que o usuário digite estas informações toda vez que desejar acessar uma funcionalidade que requeira autenticação. Uma outra possibilidade é a autenticação através de *cookies* HTTP, mas nesse caso usuários que utilizam o mesmo navegador não podem ser distinguidos um do outro, além de ser uma alternativa insegura. Um outra possibilidade, mais robusta e usada por algumas revistas eletrônicas, de autenticação através de **certificados**. Finalmente, algumas revistas e sociedades profissionais autenticam seus grupos de usuários através de domínios ou endereços IP. Alunos e pesquisadores da COPPE/UFRJ, por exemplo, têm acesso ao conteúdo de algumas revistas através de mecanismos de autenticação dessa última categoria descrita.

No discurso científico, muitas vezes o anonimato é desejável. Uma outra característica das infra-estruturas de conhecimento refere-se ao suporte a contribuições anônimas. Algumas revistas emergentes permitem que leitores façam comentários e acrescentem contribuições de forma anônima, por exemplo, a *Principia Cybernetica Web* (HARS, 2003).

A última característica que discutiremos dentro da dimensão “gerenciamento” diz respeito ao modelo de financiamento de uma infra-estrutura de conhecimento, ou seja, qual a abordagem utilizada pela revista para cobrir os custos necessários para

manter sua infra-estrutura em produção. Em outras palavras, qual o custo para manter uma infra-estrutura de conhecimento científico funcionando na Web?

Algumas abordagens foram colocadas em prática ao longo dos anos pelas revistas científicas tradicionais. Na principal delas, baseada em *inscrições anuais*, as bibliotecas de pesquisa das grandes universidades, instituições de pesquisa, e os leitores pagam uma taxa anual para acessar os artigos. Grandes revistas internacionais como ACM, Elsevier e Springer possuem esquemas desse tipo. Uma outra abordagem comum é a do *pagar por leitura* (“pay-per-view”), onde o leitor paga uma taxa toda vez que acessa um documento. O principal problema desta abordagem está em determinar um preço adequado para o artigo, que deveria levar em conta o perfil do leitor, o uso que o leitor fará do artigo (simplesmente uma navegação ou um trabalho de pesquisa em cima do mesmo), o padrão de qualidade do artigo etc. Uma terceira abordagem, que se difundiu particularmente na década de 70 (HARS, 2003), é a do *pagar por submissão*. Neste caso, são os autores dos artigos que financiam a publicação, pagando por cada artigo submetido. Se por um lado esta alternativa privilegia a submissão de artigo de melhor qualidade, por outro lado, dificulta a ampla aceitação por parte dos autores.

Alguns livros e artigos se dedicam exclusivamente à discussão dos modelos de financiamento da publicação científica. Todavia, uma discussão mais aprofundada dos aspectos econômicos relacionados ao processo de publicação científica foge ao escopo dessa monografia.

2.5.5 Tecnologia

A última dimensão a ser apresentada é a tecnologia utilizada. Nesta discussão estão incluídos: tipos de servidor utilizados na implementação das infra-estruturas de conhecimento, a tecnologia de busca empregada, as linguagens e padrões utilizados e o suporte à autenticação de documentos.

Todas as infra-estruturas de conhecimento científicas baseadas na Web utilizam algum servidor Web para retornar páginas HTML de acordo com a requisição de seus usuários. Grande parte delas também utiliza algum banco de dados para armazenar os nós de conhecimento e outras informações. A utilização de um sistema gerenciador de banco de dados possibilita maior flexibilidade e customização na exibição do conteúdo, já que as requisições do usuário podem ser transformadas em consultas ao banco, retornando diferentes apresentações de um documento, por exemplo. Uma outra

ferramenta utilizada com frequências por estes sistemas é uma máquina de busca. Em raros casos, de acordo com HARS (2003), uma máquina de workflow é também empregada no processo de revisão, para a coordenação das atividades dos editores, revisores e autores.

No tocante à tecnologia de busca empregada, a rápida evolução da Internet levou ao desenvolvimento de sofisticadas máquinas de busca. Alguns consideram que a primeira geração destas tecnologias, representada por nomes como Altavista e Yahoo, apresentava como grave problema o número gigantesco de documentos relevantes para uma determinada consulta do usuário. Isto é agravado pelo fato de que os usuários são relutantes em explorar mais do que 10 resultados numa lista (BRIN & PAGE, 1998). A tecnologia desenvolvida pelo grupo de pesquisa do Google, considerado como parte da “segunda geração” da Web, procurou minimizar de alguma forma este problema. A idéia é calcular a relevância (*page rank*) de cada página na Web, a partir da análise da rede de ponteiros entre as páginas (os “*links*”). A mesma tecnologia de busca disponível de forma gratuita na Web poderia ser empregada nos portais de conhecimento científico. Muitas infra-estruturas já exploram a busca de artigos por palavras-chave, por autores, título e categorias. Algumas também suportam a busca pelo texto completo. Uma outra possibilidade que poderia ser explorada seria uma busca baseada no comportamento do usuário.

O próximo item a ser discutido refere-se a os padrões empregados pelas infra-estruturas do conhecimento. A maioria delas utiliza o padrão HTML para a confecção de suas páginas, e empregam *JavaScript* e *DHTML* no lado cliente para proporcionar uma maior interatividade. Outras escolheram padrões como ASP, da Microsoft, ou JSP, no lado servidor, para a geração de conteúdo dinâmico. A tendência para os próximos anos é o emprego maciço da tecnologia XML, que permite a separação do conteúdo de um documento de sua apresentação.

Finalmente, o último ponto a ser discutido diz respeito à tecnologia empregada na autenticação dos documentos, um requisito importante para evitar a cópia e a transmissão do conteúdo eletrônico. Existem várias propostas de padrão para resolver este problema, mas uma discussão destas técnicas foge ao escopo deste trabalho, e pode ser encontrada em MINTZER et al. (1998).

2.6 Sumário e Conclusões

Este capítulo teve como principal objetivo realizar uma revisão ampla na literatura que versa sobre as infra-estruturas de conhecimento científico, começando pelas tradicionais, que replicam características do modelo antigo, baseado no papel, e chegando nas emergentes, que têm buscado novas formas de comunicação sem replicar necessariamente as inadequações do modelo antigo.

Neste contexto, fizemos também uma análise teórica do processo de publicação científica, com o objetivo de entender o sistema de publicação tradicional, seus problemas e limitações, e de que forma a tecnologia da informação poderia minimizar ou resolver alguns problemas importantes do sistema tradicional. Prosseguindo, apresentamos a definição de uma infra-estrutura de conhecimento científico, e realizamos uma análise empírica de tais infra-estruturas, na qual apresentamos alguns modelos emergentes que foram implementados nos últimos anos.

Finalmente, caracterizamos uma infra-estrutura do conhecimento científico seguindo o *framework* de características proposto por HARS (2003), que nos forneceu as bases para o projeto do ambiente CEJ, principal objetivo desta monografia. Nesta caracterização, abordamos o modo de interação em uma infra-estrutura de conhecimento, discutindo os aspectos de colaboração, gerência de processos e sistemas de workflow, mecanismos de percepção e sistemas de recomendação. Abordamos também as características relacionadas à apresentação dos nós de conhecimento, à estruturação de tais nós, ao gerenciamento das infra-estruturas de conhecimento e, finalmente, discutimos os aspectos tecnológicos.

O principal objetivo do capítulo seguinte será traçar uma análise comparativa de alguns sistemas de anotações baseados na *Web*, com o intuito de identificar as melhores oportunidades para o ambiente proposto nesta monografia.

Capítulo 3

Sistemas de Anotações baseados na Web

3.1 Considerações iniciais

O advento da *World Wide Web* veio a simplificar sobremaneira o processo de publicação de documentos, bem como sua busca e recuperação. Todavia, o modelo de colaboração sobre documentos na Web se mostrou *assimétrico*, com os *autores* dos documentos sendo *produtores ativos*, e os *leitores* sendo *consumidores passivos*.

Um sistema de anotação (*annotation system*) possibilita a criação e recuperação de artefatos chamados *anotações*, compondo os chamados *documentos virtuais*. Estes nada mais são do que uma fusão do documento original com as anotações compostas pelos leitores do documento. Desta forma, um sistema de anotação permite que leitores se tornem *consumidores ativos*, colaborando com os autores na produção da informação.

Vannevar Bush, em seu famoso artigo intitulado “As we may think” (BUSH, 1945), foi o primeiro a contemplar um sistema em que múltiplos usuários deixariam “pegadas” (*trails*) no espaço de um documento. O artigo, publicado em 1945 em uma seção especial da revista *The Atlantic Monthly*, antecipava o aparecimento da Web e, mais recentemente, de dezenas de ferramentas para anotações sobre documentos nela disponíveis.

Sistemas de anotações "são sistemas baseados em hipertexto que permitem que múltiplos usuários conectem novos nós à estrutura existente de um documento" (WHITTINGTON, 1996). A nova estrutura produzida deve ser visível a um grupo de usuários que poderão adicionar novos nós, criando linhas de discussão.

Na visão de DENOUE (2002), uma *anotação* pode ser vista como um par (âncora, descrição). A *âncora* codifica o posicionamento da anotação no documento anotado, podendo ser o documento como um todo, um trecho selecionado de um parágrafo etc. A *descrição* é o conteúdo embutido pelo usuário na âncora, podendo ser um comentário textual, um apontador para um outro documento, um aviso sonoro etc.

Neste capítulo traçamos uma análise comparativa de alguns dos principais sistemas de anotações na Web, com o objetivo de identificar as melhores oportunidades para o

ambiente proposto nesta monografia. O restante do capítulo está organizado da seguinte maneira. Na seção 3.2 caracterizamos um sistema de anotação através de um *framework* de 12 características proposto por GRAMLICH (1997). Na seção 3.3 apresentamos as principais arquiteturas adotadas no projeto de um sistema de anotação para a Web, de forma a permitir uma análise comparativa das ferramentas dentro da perspectiva de cada grupo. Na seção 3.4 são apresentadas as principais limitações da infra-estrutura da Web no suporte aos sistemas de anotações.

Na seção 3.5 é feita uma análise comparativa de alguns sistemas de anotação na Web, tanto comerciais quanto acadêmicos. Foram selecionadas ferramentas pertencentes a cada uma das arquiteturas descritas na seção 3.4. Um quadro comparativo das ferramentas analisadas é apresentado na seção 3.6. Finalmente, a seção 3.7 faz um sumário do capítulo e apresenta as principais conclusões.

3.2 Características dos sistemas de anotações

Segundo GRAMLICH (1997), é possível caracterizar um sistema de anotação e diferenciá-lo dos outros seguindo um conjunto de 12 características: aplicação, arquitetura, autenticação, autoria, granularidade, moderação, notificação, armazenamento, tipagem, controle de versão, visibilidade e votação.

A *aplicação* do sistema determina se o sistema é genérico ou foca numa aplicação específica. Por exemplo, alguns sistemas se aplicam à revisão de documentos eletrônicos, enquanto outros são focados na discussão no ambiente educacional.

Quanto à *arquitetura*, os sistemas podem ser baseados em um servidor, em uma aplicação *proxy* ou na extensão de um navegador. Na próxima seção serão discutidas as três abordagens. A *autenticação* define o mecanismo utilizado pelo sistema para garantir a identidade do autor de uma anotação. Já a *autoria* define como o sistema permite criar as anotações, se a anotação pode ser texto, HTML, figura, som etc. A *moderação* do sistema diz respeito à forma de divisão do controle entre o administrador e seus usuários.

A *granularidade* de um sistema de anotações pode ser *fina*, quando ele permite anotações em qualquer lugar do documento; *média*, quando permite anotações em parágrafos, seções ou unidades de anotação estabelecidas pelo sistema; ou ainda *grossa*, quando anotações são permitidas apenas ao documento como um todo. Em geral, quanto

mais fina é a granularidade do sistema, mais difícil se torna manter as anotações embutidas no documento, quando este sofre modificações.

Outra característica que ajuda a definir um sistema é a forma de *notificação* ao usuário quando surgem novas anotações. Notificações podem ser feitas através do correio eletrônico (*email*), através de índices de “mudanças recentes” embutidos em uma página do sistema etc. Alguns sistemas não utilizam nenhum mecanismo de notificação, deixando ao usuário a tarefa de verificar a existência de novos comentários e anotações.

Quanto ao *armazenamento*, as anotações podem ser guardadas no próprio documento ou fora dele. Isto em geral depende do fato de o usuário ter ou não direito a alterar o documento. A *tipagem* define se uma anotação é um comentário, uma idéia etc, sendo que alguns sistemas não utilizam nenhuma forma de tipagem. O *controle de versão* é outra característica importante, e define se o sistema trata ou não a evolução do documento e das anotações ao longo de suas diversas versões. Alguns sistemas simplesmente desconsideram o versionamento das anotações, e quando ocorre uma alteração na estrutura e/ou no conteúdo do documento, o sistema tenta exibir as anotações de forma “correta”, embora nem sempre isso seja possível. Esta situação pode acontecer muito frequentemente em alguns sistemas que praticam versionamento dos documentos, e é conhecida na literatura como o *problema das anotações órfãs*.

Quanto à *visibilidade*, sistemas de anotações podem permitir anotações privadas, públicas ou anotações do grupo. Finalmente, o item *votação* define se o sistema usa algum mecanismo de votação ou de graduação para manter uma anotação na base.

3.3 Classificação dos sistemas quanto à arquitetura

Em (BRUSH, 2002), a autora classifica as ferramentas de anotação disponíveis para a Web em três grandes grupos, baseado na sua arquitetura e implementação:

- sistemas *baseados no servidor*
- sistemas *baseados em uma aplicação proxy*
- sistemas *baseados na extensão do navegador cliente*

Nos sistemas *baseados no servidor*, é necessário carregar o documento a ser anotado em um servidor particular, e o documento e as anotações são armazenados neste servidor. O

documento ganha uma formatação especial para que possa ser anotado. São representantes típicos desta categoria os sistemas *CaMILE* e o *DocReview*.

Nos sistemas *baseados em uma aplicação proxy*, documentos são acessados através de uma aplicação intermediária (o *proxy*), que estabelece uma ponte entre o usuário e o documento. Neste caso, qualquer documento disponível na Web pode ser anotado, já que não há a obrigatoriedade de o documento ser carregado em um servidor específico. Exemplos de ferramentas desse tipo incluem o software *CritLink* e o sistema *CoNote*.

Finalmente, nos sistemas *baseados na extensão do navegador cliente*, um navegador particular (como o *Netscape* e o *Internet Explorer*) é estendido para permitir que qualquer documento Web possa ser anotado. É necessário em geral instalar um software (*plugin*) para que as anotações possam ser criadas e lidas. Exemplos desses sistemas incluem o sistema *ComMentor*, que estende o navegador *Mosaic*, e a ferramenta *Annotea*, que estende o navegador *Amaya* do W3C.

3.4 Infra-estrutura de apoio às anotações

Ao se projetar um sistema de anotação para a Web, o primeiro passo é a escolha de uma *arquitetura* para o sistema, que grosso modo pode ser uma das três arquiteturas descritas na seção anterior: baseada no servidor, baseada em uma aplicação *proxy*, ou ainda baseada na extensão de um navegador para a Web. É necessário também que se escolha uma forma de *armazenamento* para as anotações, decidindo-se se as anotações serão guardadas dentro do próprio documento, ou se serão externas a ele. Armazenar as anotações no próprio documento pode ser uma estratégia problemática, sobretudo quando as anotações são compartilhadas por diferentes usuários.

Um formato adequado para a codificação das anotações é também uma escolha importante. O advento da *Web Semântica* (BERNERS-LEE, 2001) trouxe a possibilidade de que anotações sejam armazenadas, recuperadas e trocadas em um formato de metadados como XML. A linguagem HTML, apesar de se mostrar relativamente adequada para a visualização das anotações, é bastante limitada em sua expressividade. Por outro lado, a linguagem XML é muito mais expressiva, sendo capaz de descrever tanto o conteúdo quanto a estrutura de uma anotação.

De acordo com VASUDEVAN & PALMER (1999), o projeto de um sistema de anotação pode ser dificultado em decorrência de algumas limitações da infra-estrutura atual da *World Wide Web*. Alguns exemplos de restrições deste tipo são listados a seguir:

- A extensão dos principais navegadores para a Web (*Internet Explorer, Netscape, Opera* etc) é limitada em função de projetos (e conseqüentemente de implementações) divergentes dos navegadores e em decorrência dos mecanismos de segurança existentes no momento. Um exemplo conhecido acontece quando dois frames são criados em um documento HTML. Por questões de segurança, um frame só poderá acessar o conteúdo do outro caso ambos os documentos estejam localizados no mesmo servidor de arquivos ou dentro de um mesmo domínio.
- Quando um documento HTML é requisitado e carregado em um navegador, seu conteúdo é manipulado através do modelo de objetos do documento - o DOM (*Document Object Model*). O acesso ao DOM é feito sempre através de uma linguagem de script cliente suportada pelo navegador, em geral *JavaScript* ou *VBScript*. Através do DOM é possível acessar trechos do documento onde anotações podem ser inseridas e posteriormente recuperadas. Todavia, o DOM só pode ser acessado através de scripts que estejam embutidos na página sendo carregada, e necessariamente escritos na linguagem de scripts hospedeira do navegador. Isto dificulta sobremaneira a construção de um sistema de anotação, pois impede que uma aplicação *proxy* externa ao documento tenha controle sobre o DOM. Este suporte ao DOM é conhecido como *nível 0* (W3C DOM WORKING GROUP, 1998). Existe no momento um grupo no W3C trabalhando na extensão do DOM, de forma a torná-lo acessível a programas externos ao browser. Num futuro próximo, é provável que linguagens de script do lado servidor, como *PHP, ASP* e *JSP*, também tenham disponíveis uma API para o acesso ao DOM.
- Em geral, uma anotação não é identificada de forma global, ficando associada sempre a um determinado servidor de anotações. Como fazer com que um documento aponte para uma anotação que esteja em outro documento, já que não se associa uma URL a uma anotação? É bem verdade que algumas iniciativas já foram feitas no sentido de minimizar esse problema. Em (DENOUE, 1999), por exemplo, os autores propõem codificar as anotações em um formato estendido de

URL. De qualquer maneira, uma anotação deveria ser considerada, via de regra, um documento como qualquer outro, e como tal ser acessível através de uma URL na Web.

3.5 Análise comparativa de Sistemas de Anotações na Web

Nesta seção faremos uma análise comparativa de alguns sistemas de anotação para a Web. Fizemos uma seleção criteriosa das ferramentas a serem analisadas, levando em conta as arquiteturas apresentadas na seção 3.4 deste artigo, incluindo iniciativas tanto acadêmicas quanto comerciais e, finalmente, contrastando trabalhos mais antigos com os mais recentes.

Começaremos analisando as ferramentas tipicamente baseados *no servidor*: o software *CaMILE* e o *DocReview*. Prosseguindo, descreveremos dois sistemas baseados em uma aplicação *proxy*: *CritLink* e *CoNote*. Finalmente, faremos uma análise de três sistemas baseados na extensão de um navegador Web: os acadêmicos *ComMentor* e o *Annotea*.

3.5.1 CaMILE

CaMILE (*Collaborative and Multimedia Interactive Learning Environment*) é um ambiente de colaboração para a Web aplicado ao ensino e baseado em anotações. O projeto de pesquisa foi conduzido por Mark Guzdial e financiado por diversas entidades, entre elas a agência ARPA (*Advanced Research Projects Agency*) e a fundação NSF (*The National Science Foundation*).

Neste sistema, anotações são apontadas sempre através de uma URL: qualquer documento na Web pode conter um apontador para uma linha de discussão (*thread*) do ambiente CaMILE.

Quando um usuário deseja criar uma anotação, o sistema solicita, através de perguntas, que ele identifique o tipo da anotação. Por exemplo, uma anotação poderá ser uma "pergunta", uma "idéia nova", uma "contraposição" etc. Para cada tipo de anotação, o sistema sugere uma frase inicial, que serve como ponto de partida para o usuário preencher o texto da anotação. Evidentemente, os tipos possíveis para uma anotação são configuráveis em cada instância do software, dependendo da aplicação. Anotações no sistema podem incluir tags em HTML, permitindo a inclusão de imagens, apontadores etc.

Existem duas formas de visualizar uma anotação: clicando em um dos apontadores na página inicial do sistema, ou em um apontador localizado na própria anotação. O sistema exibe o título, o autor, o tipo e o conteúdo da anotação. Linhas de discussão são refletidas pela indentação do texto das anotações, como se pode notar na figura abaixo extraída de (GUZDIAL, 1997a).

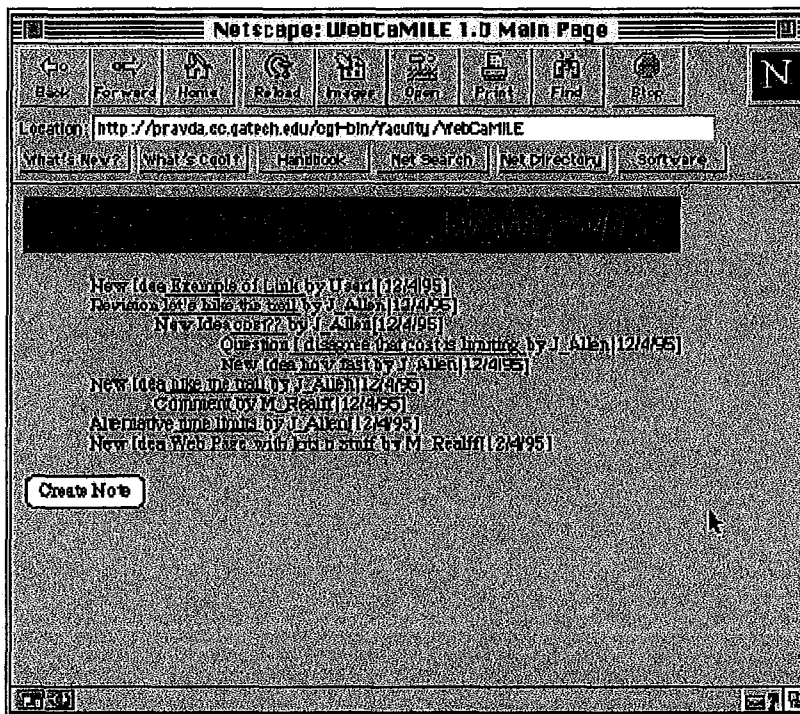


Figura 3.1 – Screenshot de uma instância do WebCaMILE

De acordo com GUZDIAL (1997a), uma diferença importante entre o sistema CaMILE e o serviço de *newsgroups* é que CaMILE suporta a chamada "colaboração através de âncoras", isto é, toda anotação pode ser referenciada de forma única através do *browser*. As âncoras funcionam como índices e como lembretes do que os estudantes devem discutir em uma determinada linha de discussão. Tipicamente, são os professores os responsáveis pela criação das âncoras.

A ferramenta CaMILE foi utilizada experimentalmente por estudantes da Alemanha em diferentes disciplinas, como Engenharia Química e Civil, Ciência da Computação e Astronomia. Recentemente o desenvolvimento do sistema CaMILE foi descontinuado, dando origem a outro projeto colaborativo, o *Swikis*.

3.5.2 DocReview

Trata-se de um sistema para a Web desenvolvido por Charles Hendricksen, aluno da universidade de Washington, com o intuito de facilitar o processo de colaboração na pesquisa científica. Sua tese de doutorado (HENDRICKSEN, 2002) apresenta o ambiente colaborativo *The Research Web* (RW), do qual a ferramenta *DocReview* faz parte, e aborda a colaboração assíncrona na pesquisa científica.

A principal aplicação deste sistema é, sem dúvida alguma, a revisão de documentos científicos. De acordo com HENDRICKSEN (1999), *DocReview* transforma os leitores de um documento em revisores, permitindo que pesquisadores possam contribuir, refinar e corrigir artigos científicos. A ferramenta é também utilizada na construção de bibliografias comentadas, de glossários especializados, e na revisão de trabalhos acadêmicos. A figura a seguir ilustra uma instância do sistema utilizada em uma universidade de *Washington*:

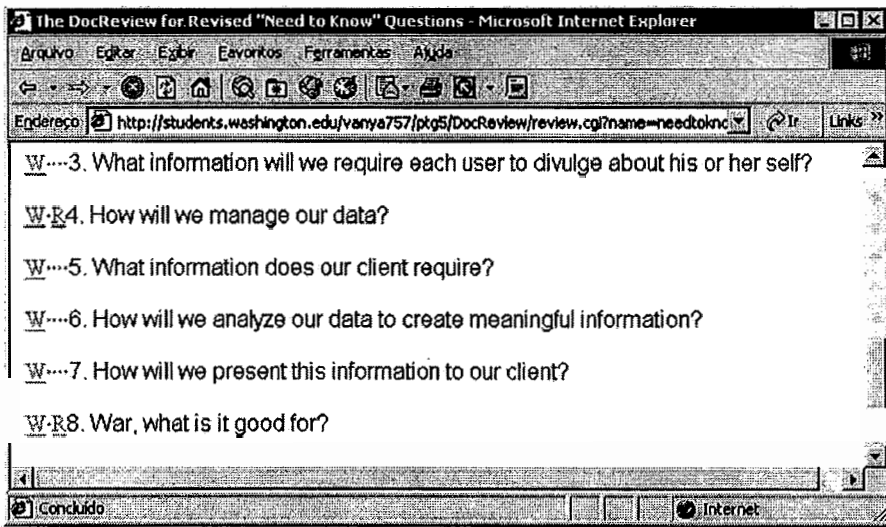


Figura 3.2 – Screenshot de uma instância do DocReview

Neste sistema, anotações são sempre adicionadas a uma seção do documento conhecida como *segmento de revisão*. Apenas ao editor do sistema é permitida a criação dos segmentos de revisão. A presença de anotações em um determinado segmento de revisão é indicada por apontadores *R* (*read*). Quando o leitor clica em um desses apontadores, o sistema exibe uma lista com as anotações do segmento. Um segmento que ainda não foi comentado simplesmente não apresenta o apontador *R*.

Revisores criam anotações clicando no apontador *W* (*write*), que aparece no início de todo segmento de revisão. Ao clicar no apontador *W*, o sistema exibe um formulário solicitando ao revisor que digite seu nome, email e conteúdo da anotação.

Aproximadamente 400 instâncias da ferramenta foram instaladas no mundo inteiro desde 1995 (HENDRICKSEN, 1999). O sistema tem sido utilizado na revisão de artigos de revistas científicas, em conferências e workshops, como é o caso da conferência sobre *Computer Supported Cooperative Learning* (CSCL), realizada em 1999.

Uma das limitações do sistema está no fato de que as anotações ficam restritas a porções do documento definidas pelo editor (os segmentos de revisão). Como pontos positivos, podemos destacar a simplicidade da ferramenta e o fato de ela poder ser baixada gratuitamente (via FTP) e utilizada por pesquisadores de todo o globo.

3.5.3 CritLink

Trata-se de uma ferramenta tipicamente baseada em uma aplicação *proxy*, desenvolvido por *Ka-Ping Yee* no instituto *Foresight*. Ele é responsável pela confecção do termo *mediador*, o qual ele prefere ao termo *proxy*. De forma resumida, um mediador, uma vez acionado por um navegador cliente, faz requisições a outros servidores Web, adicionando informação aos documentos recebidos (YEE, 1999). Desta forma, um mediador proporciona ao usuário uma visão mais rica do documento. No caso do sistema CritLink, o mediador adiciona apontadores e anotações aos documentos requisitados pelo cliente.

Para acessar o sistema, o usuário deve navegar para a página inicial do CritLink e digitar a URL da página que deseja visitar. O mediador recupera esta página, busca as anotações do documento no servidor de anotações, e retorna a mesma para o usuário. A partir deste momento, toda a navegação realizada pelo usuário (através de apontadores) passa a ser processada pelo mediador.

Anotações podem aparecer em duas formas distintas: adicionadas no final do documento ou embutidas no próprio texto. No segundo caso, são exibidas através de "setas" que circundam o texto anotado. As setas nada mais são do que apontadores para o comentário.

A granularidade das anotações pode ser considerada *fina*, já que podem ser adicionadas a qualquer ponto do documento. O único problema é a interface pouco amigável: é necessário digitar as linhas do documento que se deseja anotar numa janela de comentário!

Outra crítica está na apresentação das anotações. Há uma sobrecarga de informações que são exibidas quando se navega para uma anotação, como o email do autor, o endereço da página anotada etc.

3.5.4 CoNote

Trata-se de um projeto desenvolvido na "Cornell University" baseado no sistema HyperNews e escrito em Perl (DAVIS, 1996). No sistema *CoNote*, as anotações ficam restritas a pontos escolhidos pelo autor no momento da criação do documento. Uma das vantagens desta abordagem é que as anotações permanecem no documento quando uma nova versão é produzida. A desvantagem é que as anotações ficam limitadas a esses pontos pré-definidos, dando pouca liberdade ao usuário do sistema.

Para cada anotação criada no documento, são exibidos o nome do autor, a data de criação, e um apontador para o texto da anotação.

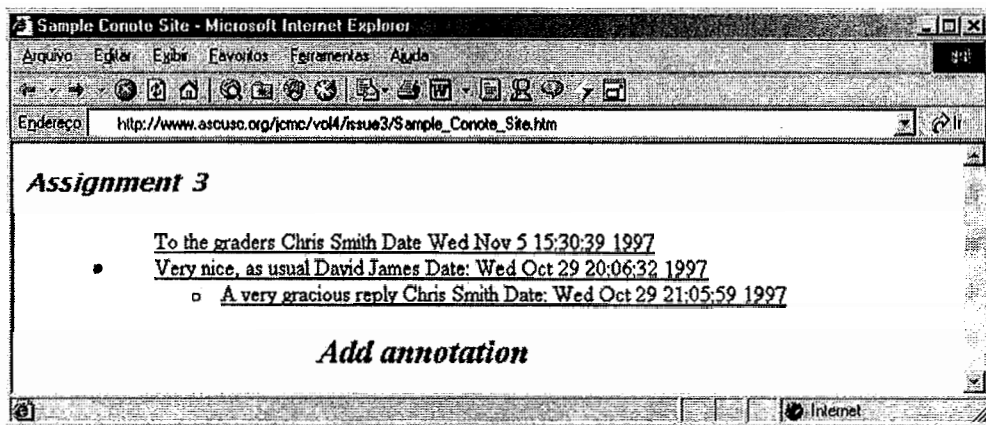


Figura 3.3 – Screenshot de uma instância do CoNote

A lista das anotações em um determinado ponto de anotação é ordenada por data. Uma anotação feita sobre outra anotação é indentada, de forma semelhante a uma linha de discussão, conforme se pode notar na figura anterior.

Para adicionar uma anotação, basta clicar no botão "Reply" localizado nos pontos definidos para as anotações. Se já existir uma lista de anotações em um determinado ponto, o botão de "Reply" aparece no final.

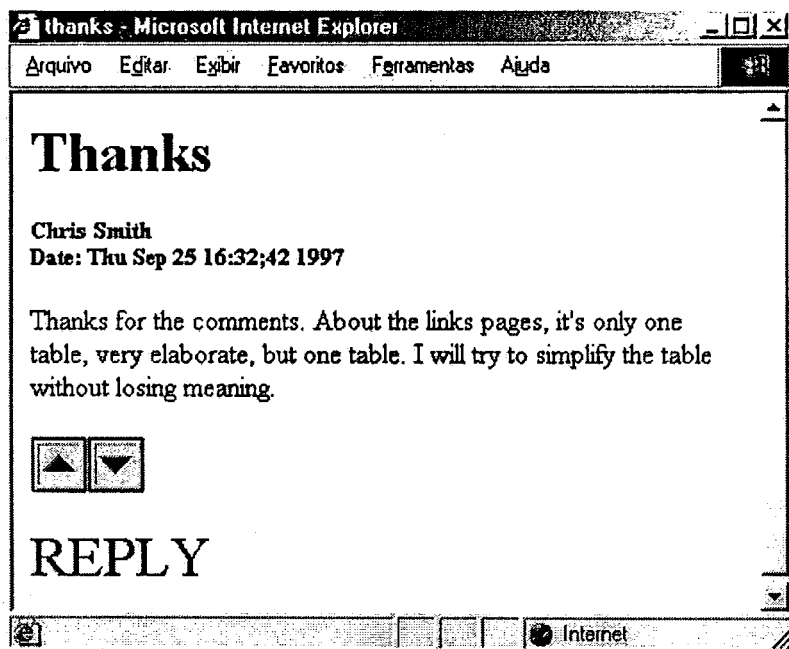


Figura 3.4 – Screenshot de uma instância do CoNote

O acesso às anotações no sistema pode ser dividido nos seguintes papéis:

- *viewer* - pode ler o documento, mas não as anotações
- *reader* - pode ler o documento e as anotações
- *user* - pode ler e adicionar anotações
- *author* - pode ler, adicionar e deletar anotações

É possível também que leitores anônimos, não cadastrados em nenhum grupo no *CoNote*, acessem determinados documentos do sistema, bastando para isto que o grupo defina um papel *default*. Quanto à visibilidade, o sistema suporta apenas anotações do grupo.

Finalmente, o sistema suporta a busca de anotações por documento, autor e data.

3.5.5 ComMentor

ComMentor (RÖSCHEISEN et al., 1995) é um exemplo típico de sistema de anotações baseado na extensão do navegador cliente. Foi escrito em Perl e desenvolvido sob a égide do projeto de bibliotecas digitais integradas da universidade de Stanford.

Neste sistema, o navegador *Mosaic* foi modificado para suportar funções adicionais. Exemplos dessas funcionalidades incluem itens de menu para a criação de anotações, anotações embutidas no próprio texto, janelas para a visualização do conteúdo da anotação etc.

A arquitetura do sistema baseia-se no conceito de “conjuntos de anotações” (*annotation sets*). Toda a notação no sistema deve pertencer a um determinado conjunto, anotando uma determinada porção de um determinado documento. São três as possibilidades de conjunto de anotações: anotações *privadas*, *públicas*, ou *do grupo*. Um grupo pode ter a ele associado mais de um conjunto de anotações, cada um utilizado com um objetivo diferente.

Para adicionar uma anotação, o leitor seleciona, no documento HTML, a porção do texto que deseja anotar e escolhe a opção *anotar* no menu do navegador. No momento em que é criada uma anotação, o sistema solicita, através de uma caixa de diálogo, que o usuário digite o nome e o conteúdo da anotação. O usuário deve informar também a que conjunto a anotação pertence. A figura 3.5 ilustra uma das telas do sistema:

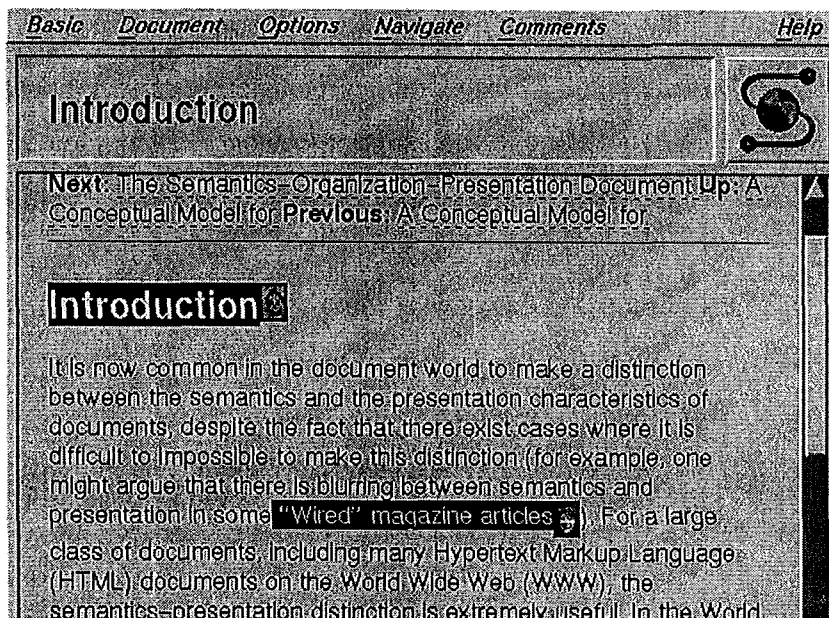


Figura 3.5 – Screenshot de uma instância do ComMentor

Como se pode notar, comentários são identificados por ícones, representando diferentes usuários do sistema. Exemplos de telas ilustrando as principais funcionalidades do sistema *ComMentor* podem ser encontradas em <http://www-diglib.stanford.edu/rmr/TR/shots>.

No sistema *ComMentor*, anotações podem também ser comentadas, já que são tratadas como documentos. Ao navegar por uma anotação, as anotações sobre ela feitas são exibidas na mesma janela, de forma semelhante a uma linha de discussão. O único inconveniente é que o leitor só consegue verificar se uma anotação foi comentada navegando sob a mesma.

No sistema *ComMentor*, anotações podem também ser comentadas, já que são tratadas como documentos. Ao navegar por uma anotação, as anotações sobre ela feitas são exibidas na mesma janela, de forma semelhante a uma linha de discussão. O único inconveniente é que o leitor só consegue verificar se uma anotação foi comentada navegando sob a mesma. A figura a seguir ilustra uma tela do sistema:

Recentemente, o desenvolvimento do sistema foi descontinuado, e de acordo com os próprios criadores, a ferramenta não é mais compatível com as versões atuais dos principais sistemas operacionais e navegadores para a Web.

3.5.6 Annotea

Annotea é um projeto da *World Wide Web Consortium* (W3C) com o objetivo de incentivar a experimentação das recentes especificações para a Web, como XML, XHTML, RDF, MathML, XPointer, entre outras. Sua arquitetura é baseada na extensão do navegador oficial do W3C - o Amaya.

Trata-se fundamentalmente de um sistema de anotações *centrado no documento*, onde o usuário navega em um documento e eventualmente verifica suas anotações (KAHAN et al., 2002). Existe uma outra classe de sistemas de anotações, *centrados na discussão*, onde o usuário predominantemente trabalha sobre linhas de discussão e eventualmente navega para um documento apontado por uma das mensagens da discussão. Exemplo típico de sistema pertencente a esta categoria é o *HyperNews*.

Navegando no Amaya, o leitor pode comentar um documento XHTML de diversas maneiras: anotando um trecho selecionado do documento, anotando um elemento específico do documento, como um parágrafo, ou anotando o documento como um todo.

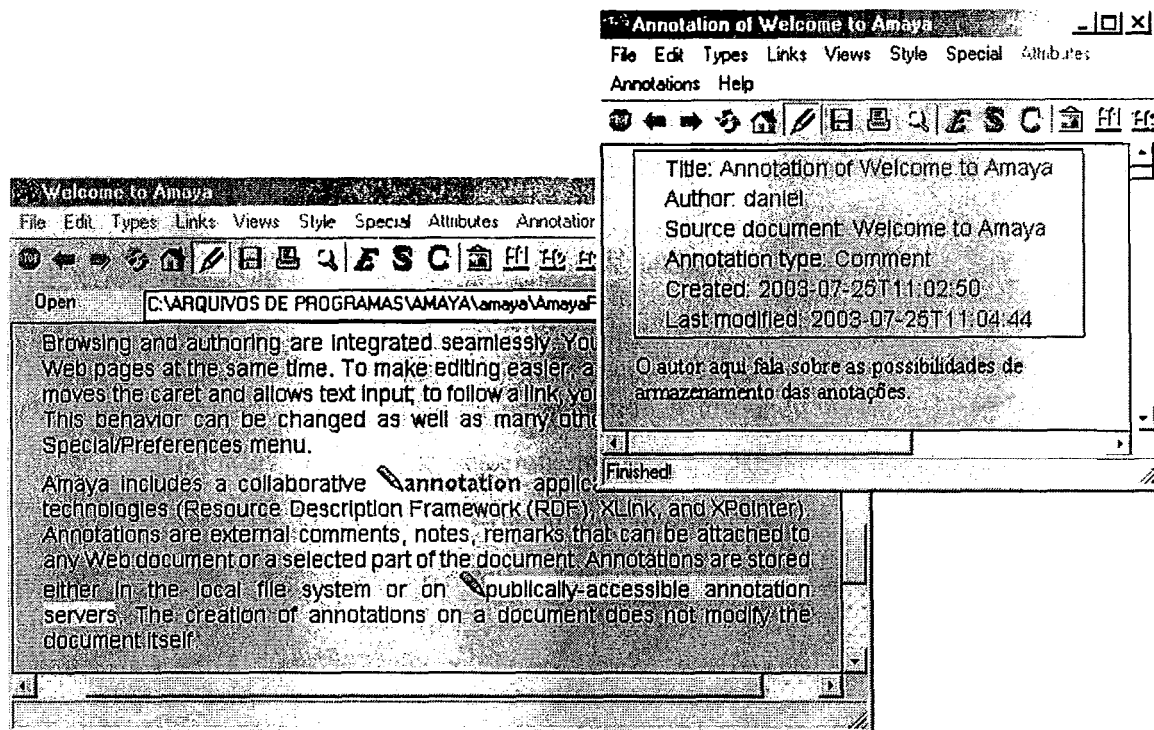


Figura 3.6 – Screenshot de uma instância do Annotea

O corpo da anotação nada mais é do que um documento XHTML, e como tal pode ser editado no navegador Amaya como qualquer outro documento. A janela de exibição de uma anotação mostra, além do corpo da anotação, meta-informações da anotação, como autor, data de criação etc.

O sistema suporta anotações *privadas* e *compartilhadas*. No momento de criação de um anotação, ela é considerada privada e local, sendo armazenada no diretório do Amaya. O usuário pode posteriormente especificar um servidor de anotações remoto e decidir salvar a anotação neste servidor. Neste caso, a versão local da anotação é apagada e uma nova versão remota e compartilhada é criada no servidor. Toda modificação subsequente desta anotação passa, então, a ser realizada no servidor remoto.

Anotações são exibidas no documento através do chamado *elemento-A*, que é um XML contendo um XLink apontando para o corpo da anotação. Este elemento é exibido no

documento através de um ícone especial (a figura de um lápis) e, quando o usuário clica sobre ele, o sistema enfatiza o alvo da anotação, que pode ser um trecho selecionado, um parágrafo ou mesmo o documento completo. Para navegar sobre a anotação, o usuário deve dar um duplo clique sobre o elemento-A, o que fará com que o sistema abra uma janela exibindo tanto a meta-informação quanto o corpo da anotação.

Para responder a uma anotação, o usuário pode modificar o seu corpo ou anotá-la. Desta forma, uma das limitações do sistema está no fato de que, para acompanhar uma linha de discussão, o usuário vai precisar navegar por cada uma das anotações.

3.6 Quadro comparativo das ferramentas analisadas

A tabela a seguir apresenta um quadro comparativo das ferramentas analisadas nesta pesquisa. Nesta comparação, foram incluídas as seguintes características: *aplicação* da ferramenta, tipo de *arquitetura*, formas de *autoria* das anotações, *granularidade* das anotações, mecanismos de *notificação* suportados pela ferramenta, forma de *armazenamento* das anotações, *tipagem*, suporte ao *controle de versão*, *visibilidade*, suporte a *linhas de discussão*, suporte a mecanismos de *busca de anotações* e, finalmente, o *status* corrente do sistema.

Tabela 3.1 – Comparação das ferramentas de anotações para a Web

<i>Ferramenta</i> <i>Característica</i>	CaMILE	DocReview	CritLink	CoNote	Commentor	Annotea
Aplicação	Ensino	Colaboração na pesquisa científica	Genérico	Genérico	Genérico	Genérico
Arquitetura	Servidor	Servidor	Proxy (mediador)	Proxy	Extensão do Mosaic	Extensão do Amaya
Autoria	HTML, Figura, link etc.	Texto	Texto	Texto	HTML	XHTML
Granularidade	Grossa	Média (segmentos de revisão)	Fina	Média (pontos definidos pelo autor)	Fina	Fina
Notificação	N/D	N/D	Não	N/D	N/D	N/D
Armazenamento	N/D	N/D	Externa ao documento (servidor de anotações)	Externa	Externa (HTML)	Externa (RDF e XHTML)
Tipagem	Idéia, pergunta etc (pode ser configurada)	Não	N/D	Não	Sim (anotação, comentário etc)	Não
Controle de versão	N/D	N/D	N/D	Não	N/D	Não

Visibilidade	Do grupo	Do grupo	N/D	Do grupo	Privadas, públicas e do grupo	Privadas e compartilhadas
Linhas de discussão	Sim	N/D	Sim	Sim	Sim (indentadas)	Sim
Busca de anotações	N/D	N/D	N/D	Sim (por autor, data e doc)	Sim (por autor, data e conteúdo)	N/D
Status	Descontinuado	Descontinuado	Ativo	Ativo	Descontinuado	Ativo

Legenda:

N/D: Informação não disponível.

3.7 Sumário e conclusões

Neste capítulo procuramos definir e caracterizar um Sistema de Anotação no ambiente Web, apresentando as principais limitações da infra-estrutura da Web no suporte à tecnologia de anotações, e finalmente realizamos uma revisão bibliográfica de algumas ferramentas para anotações sobre documentos na Web.

Apesar de existirem inúmeras iniciativas de sistemas permitindo anotações sobre documentos na Web, não existe nenhum estudo apontando quais ferramentas seriam mais apropriadas para quais aplicações. Isto dificultou sobremaneira a identificação da melhor arquitetura a ser adotada para o contexto de um ambiente de publicação científica. Alguns problemas parecem ser intrínsecos à maioria das ferramentas analisadas, e merecem uma atenção especial.

Primeiro, deve-se salientar o fato de que cada uma das ferramentas analisadas utiliza um formato próprio para o armazenamento das anotações, bem como uma API para a busca, a escrita e a recuperação das mesmas. Isto torna difícil a construção de sistemas de busca e indexação das anotações. Nesse sentido, a iniciativa mais promissora dentre as ferramentas analisadas é, sem dúvida alguma, a da ferramenta *Annotea*, na qual são utilizados padrões abertos como RDF, XPointer, XHTML e outros. Todavia, o sistema só funciona no momento com o navegador oficial da W3C – o *Amaya* - não sendo compatível com outros navegadores populares, como é o caso do *Internet Explorer* e do *Netscape Navigator*.

Uma outra ineficiência de alguns sistemas é que as anotações em geral são armazenadas em um servidor de anotações, o que é ineficiente quando as anotações são do tipo privadas. Uma solução interessante foi implementada no sistema *Annotea*, em que as

anotações privadas ficam armazenadas na estação local do usuário, enquanto que as compartilhadas são enviadas a um servidor remoto de anotações.

Podemos notar claramente a carência de mecanismos de notificação dos usuários, quando surgem novas anotações sobre um determinado documento. Uma abordagem simples seria o sistema enviar automaticamente um email aos usuários do grupo, toda vez que surgisse uma nova anotação no documento compartilhado. Todavia, isto poderia ocasionar uma sobrecarga de mensagens caso o número de anotações crescesse em demasia. Uma solução alternativa seria exibir um índice com as últimas anotações do documento em uma determinada página, onde seriam também exibidas outras meta-informações referentes ao documento.

Um outro ponto a ser destacado nas ferramentas analisadas é a carência de informações a respeito do versionamento dos documentos. A grande maioria das ferramentas simplesmente não trata o controle de versões, isto é, quando uma nova versão do documento é gerada, o sistema tenta apontar as anotações criadas no contexto da versão antiga para o novo documento. Se isso não é possível, a estratégia mais frequente é exibir as anotações *órfãs* no final do documento.

Conforme visto na seção 3.5, que fala sobre a infra-estrutura da Web de apoio aos sistemas de anotações, uma das principais limitações de tal infra-estrutura está na ausência de suporte do lado servidor ao modelo de objetos de um documento – o DOM. Existe no momento um grupo no W3C trabalhando na extensão do DOM, de forma a torná-lo acessível a programas externos ao browser. É muito provável que a próxima geração dos navegadores para a Web implementem o nível 1 do DOM, tornando possível o acesso do modelo de objetos do documento a aplicações externas.

No próximo capítulo será apresentado o CEJ – o ambiente proposto neste trabalho para a publicação científica na Web, do qual o componente de navegação sobre documentos e anotações é um dos alicerces.

Capítulo 4

O ambiente CEJ

4.1 – Considerações iniciais

A principal proposta do ambiente CEJ é fornecer mecanismos para a definição e execução de processos de publicação customizáveis, permitindo a implementação de revistas eletrônicas flexíveis. Este ambiente foi concebido dentro do contexto deste trabalho de pesquisa para possibilitar a experimentação de novos modelos de publicação baseados na mídia eletrônica e na Internet.

Neste capítulo serão apresentadas as principais propostas desta tecnologia como solução ou melhoria de alguns problemas centrais identificados no processo tradicional de publicação científica. Em seguida, serão apresentados os principais resultados da Análise e do Projeto deste ambiente. Finalmente, será descrita em linhas gerais a implementação da primeira versão experimental do ambiente CEJ.

4.2 – Propostas do ambiente CEJ

Este ambiente foi concebido inicialmente para a publicação científica em particular, mas nada impede que ele também seja experimentado com outras formas ou modalidades de publicação. As propostas serão aqui apresentadas dentro do contexto dos cinco grandes grupos de características proposto por HARS (2003) para as infra-estruturas do conhecimento – *modo de interação, estruturação do conhecimento, apresentação, gerenciamento e tecnologia* - apresentados no capítulo anterior.

4.2.1 – Modo de interação

- **Gerenciamento dinâmico de processos de publicação** – Um dos principais alicerces do ambiente CEJ é o suporte à configuração dinâmica de processos de publicação, através do emprego da tecnologia de *workflows*. Isto significa que o editor de uma revista poderá definir quais são as fases que irão compor o processo de

publicação, que atores irão participar de cada fase e como, qual o nível de interação pretendido etc. Tradicionalmente as revistas em papel ou eletrônicas colocam o foco desta interação no leitor. A proposta do ambiente CEJ é oferecer suporte à interação também com outros participantes, incluindo os autores e revisores, a través de um componente de *workflow customizável*. Esta tecnologia deve ser utilizada na automatização das diversas atividades de um processo de publicação, incluindo a composição de documentos, a submissão, revisões e comentários pelos pares, a aprovação dos artefatos produzidos nas fases anteriores, e finalmente a publicação.

- **Interações pela Web** – A proposta do ambiente CEJ é que idealmente 100% do processo de publicação seja executado e acompanhado pela Web, através de um navegador simples como o *Internet Explorer* ou o *Netscape*.
- **Suporte à autoria de documentos** – O ambiente CEJ possibilita a composição de documentos - no formato *XML* – com o auxílio de uma interface simples, acessível através de um navegador *Web*. A ausência de suporte à composição de documentos pelas atuais infra-estruturas de conhecimento é uma lacuna a ser preenchida pelas futuras infra-estruturas, de acordo com HARS (2003). Neste sentido, o sistema tenta preencher parcialmente esta lacuna ao possibilitar a edição de um documento no formato *HTML* bem como sua conversão para o formato *XML*, de forma interativa. O formato de hipertexto foi escolhido como "origem" já que as principais ferramentas de edição de texto, como o *Microsoft Word*, suportam a conversão para *HTML*.
- **Submissão de artigos pela Web** – No ambiente CEJ, a submissão de um artigo é vista como uma interação simples entre o membro da revista que vai submeter o documento (o autor), e o sistema de informação que representa a “revista”. Esta submissão é feita, da mesma forma como as outras interações, através de um navegador *Web*, via protocolo *HTTP*. Os autores têm a opção de submeter o texto do documento para ser armazenado no servidor do CEJ, ou um apontador (*URL*) para o local onde este documento esteja armazenado na Internet. Em ambos os casos o documento deve estar no formato *XML*. Conforme defendido por HARS (2003), a

localização física de um artigo na *Web* é irrelevante, e ponteiros para os artigos são preferíveis.

- **Controle da navegação sobre os nós de um artigo** - Na leitura de um documento, o sistema deverá recuperar somente as partes que serão lidas naquele momento (as duas primeiras seções do documento, por exemplo). Deve ser guardada também a informação de que nós do documento foram já lidos e que nós não foram lidos ainda. Em determinadas etapas ou atividades de um processo de publicação, membros da revista poderão acompanhar e “perceber” o andamento da leitura ou revisão de um documento por parte de outros membros.
- **Revisão pelos pares configurável** – Graças aos mecanismos de definição e execução de *workflows* customizáveis existentes no ambiente CEJ, é possível configurar diversos tipos de revisão, com o intuito de validar um artigo. Revisões poderão ser abertas, fechadas ou duplamente fechadas, e isto deve ser configurado no *template* de definição das tarefas do *workflow*. Pode-se definir mais de uma fase de revisão, seguindo a filosofia da revisão continuada proposta por ODLYZKO (2003). Finalmente, mecanismos deverão ser criados visando suportar de forma flexível a revisão *multidimensional*, onde a divisão binária entre o que foi publicado e o que não foi publicado pode ser substituída por um espectro.
- **Suporte ao comentário pelos pares** – O sistema deverá permitir também que fases de revisão sejam antecedidas, sucedidas ou executadas em paralelo com fases de *comentários pelos pares*, onde leitores são convidados a contribuir com idéias, sugestões de melhoria ou mesmo manifestações contrárias a respeito do conteúdo de um documento.
- **Suporte a anotações colaborativas de diversos níveis de granularidade** – Um dos componentes da arquitetura do CEJ, que será descrito na seção 4.4.1, oferece suporte a *anotações localizadas e colaborativas*, permitindo que participantes de um processo de publicação possam anotar elementos de granularidade fina do documento, como uma seção, um parágrafo, uma figura ou um trecho do documento.

Esta funcionalidade poderá ser utilizada em diversas atividades do ciclo de publicação, incluindo a leitura, a revisão e comentários pelos pares. O suporte a níveis finos de granularidade das anotações possibilita estabelecer um contexto para as contribuições, com o potencial de produzir revisões mais precisas e comentários mais específicos, conforme comentado no capítulo 3 desta monografia.

- **Suporte flexível a mecanismos de percepção (*awareness*)** – O sistema deverá permitir que, em determinadas atividades de um processo de publicação, um membro participante acompanhe o andamento da atividade em que ele está trabalhando, “enxergando” o que os outros participantes já produziram e no que estão trabalhando. A configuração destes mecanismos deve ser flexível e controlada, de tal forma que o editor poderá definir em que fases poderá ocorrer, e que usuário, grupos de usuários ou papéis poderão enxergar o trabalho dos outros.
- **Notificações customizadas** – O sistema deve permitir que membros de uma revista possam configurar os eventos dos quais desejam ser alertados: quando da submissão de um novo documento (de seu interesse), quando um leitor ou revisor acrescenta um comentário ou anotação sobre um documento de sua autoria ou de seu interesse, quando novos membros são associados ao seu grupo de trabalho na revista etc. Esta configuração deverá ser também flexível, isto é, deve ser possível configurar não apenas os eventos de notificação, como a frequência das notificações e até mesmo a apresentação das mensagens de notificação de forma customizada.
- **Controle da duração das atividades de um processo de publicação** – Deverão ser criados mecanismos e interfaces com o objetivo de medir a duração das atividades de um processo de publicação, como o tempo que um pesquisador leva em média para ler e para revisar um artigo.
- **Suporte à publicação instantânea e reversível** – O ambiente deverá suportar a publicação dos documentos tão logo eles sejam submetidos a uma determinada revista. Isto implica em associá-los a marcas ou *estampas*, com o objetivo de esclarecer aos leitores que determinados documentos ainda não foram alvo de uma

revisão pelos pares, por exemplo. O outro aspecto a ser considerado é que assim como os documentos poderão ser publicados de forma “instantânea”, também poderão ser removidos da revista – conforme será discutido na seção 4.4.3 – em função de a publicação ser reversível no ambiente eletrônico.

4.2.2 – Estruturação do conhecimento

- **Organização das publicações em categorias flexíveis** – O ambiente CEJ permitirá a classificação das publicações de uma determinada revista em categorias flexíveis, que agrupam artigos de um mesmo assunto ou linha de pesquisa. Apesar de os conceitos de *edição* e *volume* eventualmente caírem em desuso, o sistema deverá dar suporte a estes conceitos caso seja desejo do editor utilizá-los na sua revista.
- **Suporte a diferentes tipos de conteúdo e à interatividade** – O ambiente CEJ foi projetado para também suportar tipos de dados não textuais, como figuras e arquivos sonoros, que são naturalmente suportados pela infra-estrutura da *Web*.
- **Referências bidirecionais** – Dado um documento publicado no ambiente CEJ, o sistema deverá apresentar não apenas a lista dos artigos referenciados por este documento, como a lista dos artigos que referenciam o mesmo, e que estejam publicados na revista. Esta funcionalidade permite, por exemplo, seguir a evolução de uma idéia, além de tornar evidente e transparente o número de vezes que um documento é citado. De acordo com HARS (2003), referências bidirecionais reduzem a redundância da pesquisa científica e podem aumentar a colaboração entre pesquisadores. Vale ressaltar ainda que os artigos referenciados por um dado documento, mesmo que não estejam publicados na revista, são apresentados na lista de referências através da exibição de seus metadados, como título, nome dos autores etc, como já acontece com a maioria das revistas científicas. Finalmente, o sistema deve exibir apontadores (*links*) destas referências para suas versões em *HTML* ou *PDF*, vasculhadas na *Web* através de motores de busca como o *Google* ou o *Research Index*.

- **Armazenamento flexível dos documentos** – Conforme comentado na seção 4.2.1, o ambiente CEJ permite o armazenamento dos documentos em seu próprio servidor ou em servidores remotos. No segundo caso, o sistema armazena um ponteiro (*URL*) para o local onde o documento se encontra armazenado.

4.2.3 – Apresentação

- **Separação do conteúdo da apresentação** – Uma das propostas centrais do projeto CEJ é a separação do conteúdo de um artigo de sua apresentação. Como veremos nos itens a seguir, tecnicamente isto é possível através do emprego da tecnologia *XML* e do uso de *templates* de exibição.
- **Exibições não-lineares do mesmo conteúdo** – No ambiente CEJ, um artigo é representado como uma rede de nós associados através de arestas. Desta forma, o autor pode escrevê-lo de forma a proporcionar diferentes leituras e diferentes caminhos de navegação. Na navegação de um leitor sobre um documento, o sistema deve guardar a informação de que nós foram percorridos e de que nós não foram visitados ainda, conforme comentado na seção 4.2.1. Isto é um diferencial importante com relação aos portais das revistas eletrônicas emergentes, já que a atual infra-estrutura da Web, baseada no hipertexto, apenas registra a informação dos nós já visitados.
- **Geração dinâmica de exibições do conteúdo** – No ambiente CEJ, o mesmo conteúdo de um artigo, na forma de texto, figuras e outros dados, poderá ser exibido com diferentes formatações, de acordo com a audiência considerada, bastando para isso a definição e a aplicação de um *template de exibição*, comentado anteriormente. Isto pode colocar um desafio adicional aos autores, que precisam investir um esforço adicional na flexibilização do conteúdo, de forma a suportar uma gama maior de apresentações de acordo com o perfil do usuário leitor. Também é necessário que se criem mecanismos flexíveis para a configuração destes *templates* – por exemplo, através de *wizards*, que permitiriam que usuários sem conhecimento técnico destes *templates* pudessem configurá-los. Tecnicamente, a geração dinâmica e customizada

do conteúdo é possível através do *merge* do conteúdo do documento com algum *template* de exibição.

- **Exibições do conteúdo sensíveis ao contexto** – Isto significa que a apresentação do artigo também poderá variar de acordo com o contexto. Por exemplo, em uma fase de revisão, alguns elementos do artigo seriam destacados ou descritos em maior detalhe, enquanto que numa navegação casual sobre o documento, alguns elementos seriam exibidos e outros escondidos, de acordo com o perfil do leitor.
- **Customização da interface do usuário** – Qualquer membro de uma revista no ambiente CEJ poderá configurar a forma de exibição das páginas e do conteúdo da infra-estrutura, além da definição dos filtros a serem aplicados para a busca da informação. Essas definições são guardadas em um perfil associado ao usuário, o que na prática já acontece em alguns sistemas como, por exemplo, o *Ontolingua Server* e o *Information Systems Cybrarium* (HARS, 2003).
- **Internacionalização** – Finalmente, o ambiente CEJ deve permitir a exibição do conteúdo, sua busca e navegação em uma das línguas familiares ao leitor ou usuário da revista eletrônica.

4.2.4 – Gerenciamento de uma infra-estrutura

- **Seletividade e autenticação** – O ambiente CEJ deverá permitir a configuração flexível da audiência que poderá acessar e participar de cada funcionalidade oferecida por uma revista. Mecanismos de autenticação incluirão: 1) a criação de um nome de usuário (*login*) e uma senha para cada usuário; 2) A definição de papéis e grupos de trabalho; 3) A associação dos papéis e grupos de trabalho com os usuários. Desta forma, o acesso deverá ocorrer de forma seletiva e em alguns casos poderá ser exigida a autenticação para o acesso a determinados serviços oferecidos por uma revista.
- **Comentários e contribuições** – O ambiente CEJ permite que contribuições sejam adicionadas aos documentos, tanto na forma de anotações localizadas quanto na

forma de comentários sobre o documento como um todo, nos modos privado, público ou do grupo. No primeiro caso, as contribuições só ficam visíveis ao próprio membro criador do comentário, enquanto no segundo caso elas ficam visíveis para todos os membros. No terceiro e último caso, o acesso a uma determinada contribuição só deverá ser permitido aos membros participantes de um determinado grupo de trabalho.

4.2.5 – Tecnologia

- **Suporte à tecnologia de workflow** – Um dos componentes centrais do ambiente CEJ, conforme sugerido na seção 4.2.1, suportará o controle das atividades de um processo de publicação. Este componente para o gerenciamento de *workflows* customizáveis, que será descrito na seção 4.4.4, permitirá tanto a definição quanto a execução de processos de publicação.
- **Documentos nos formatos XML, XHTML e PDF** – Documentos no ambiente CEJ poderão ser apresentados nos formatos XHTML, XML e PDF. O formato padrão de exibição deverá ser o XHTML, que consideramos o mais adequado atualmente à apresentação do documento original em um navegador *Web*. Além disso, este formato permite a navegação sobre os objetos de um documento através do DOM, fundamental para a criação e o acesso às anotações localizadas. Por outro lado, deve também ser possível visualizar o documento no formato em que foi submetido – XML. Finalmente, o sistema deverá suportar a geração de uma versão PDF do documento, para fins de impressão.
- **Documentos são submetidos no formato XML e armazenados de forma estruturada no repositório de dados** – O sistema deverá aceitar somente documentos submetidos no formato XML, seguindo o DTD padrão exigido pela revista. Os nós do documento submetido deverão ser armazenados de forma estruturada no banco de dados, para permitir sua busca e recuperação.
- **Sistema de anotações baseado em *proxy* e centrado no documento** – Conforme discutido no capítulo 3 desta monografia, nos sistemas *baseados em uma aplicação*

proxy, documentos são acessados através de uma aplicação intermediária (o *proxy*), que estabelece uma ponte entre o usuário e o documento. Desta forma, qualquer documento disponível na *Web* poderá ser anotado, já que não há a obrigatoriedade de o documento ser carregado em um servidor específico. Uma conexão de acesso a Internet e um navegador *Web* deverão ser os únicos requisitos para acessar documentos numa revista implementada no ambiente CEJ. As abordagens *baseadas em um servidor*, e *baseadas na extensão de um navegador*, também discutidas no capítulo 3, foram desconsideradas para garantir a independência da localização dos documentos a serem anotados (no caso da abordagem baseada em um servidor) e para garantir a compatibilidade com os principais navegadores disponíveis (caso da abordagem baseada na extensão de um navegador). A abordagem escolhida oferecerá uma maior flexibilidade tanto para o leitor - que vai acessar o ambiente através de um navegador comum - quanto do autor que submeteu o artigo - que poderá armazenar o artigo em qualquer servidor público. Finalmente, a abordagem do sistema de anotações do ambiente CEJ será *centrada no documento*, o que será discutido na seção 4.4.1.

- **Anotações em XML e âncoras no formato W3C XPointer** – No ambiente CEJ, anotações são armazenadas no banco de dados, como acontece com os demais objetos do sistema, mas deve ser possível converter as anotações para o formato XML. Isto permitirá que outras aplicações “entendam” tanto os documentos - que são submetidos no formato XML - quanto as anotações do documento (também em XML). Para o posicionamento correto das anotações localizadas em um dado documento, escolhemos o padrão XPointer da W3C para a codificação das âncoras. Desta forma, optamos por uma codificação semelhante à adotada pelo sistema *Annotea* da W3C, discutido no capítulo anterior, e baseada em padrões abertos.
- **Busca de documentos e anotações por palavras-chave** – O ambiente CEJ deverá oferecer um serviço de busca por palavras-chave ou expressões em um contexto específico do documento, por exemplo, o contexto dos parágrafos, do resumo do documento, do título etc. Da mesma forma, o ambiente deverá permitir a busca das

anotações armazenadas em seu repositório baseada nos atributos de uma anotação, como título, corpo da anotação, autor etc.

- **Templates de definição de processos de publicação codificados em documentos XML** – O sistema deve ser capaz de traduzir a definição de um processo de publicação para o formato XML, assim como interpretar um *template* codificado em XML e configurar automaticamente um novo processo a partir deste documento. Desta forma, deve ser possível que duas revistas troquem definições de processos de publicação: uma revista implementada em um servidor CEJ poderia implementar um mesmo processo de publicação já adotado por outra revista, rodando em outro servidor CEJ, bastando para isso o intercâmbio destes *templates*.
- **Templates de exibição de um documento codificados em documentos XML** – O sistema também deverá permitir o intercâmbio de *templates* de exibição de documentos no formato XML, de forma análoga aos *templates* de definição de processos.

4.3 – Análise e Projeto do ambiente CEJ

Nesta seção faremos uma descrição das principais metodologias e dos padrões empregados na análise e no projeto do ambiente CEJ, principal produto desta pesquisa. Serão também apresentados alguns dos resultados destas duas fases da construção deste ambiente. É importante destacar que a análise, o projeto e a implementação deste ambiente são considerados abertos, possibilitando a extensão deste trabalho por outros grupos de pesquisa no futuro.

4.3.1 – O paradigma orientado a objetos

Na construção do ambiente CEJ, decidimos pela adoção do paradigma orientado a objetos, o qual aplicamos na análise, no projeto e na implementação deste ambiente. Esta escolha justifica-se pela possibilidade de construção flexível do ambiente através de componentes reutilizáveis – as classes. Uma outra vantagem do paradigma orientado a objetos é a aproximação do problema no mundo real de sua solução no mundo virtual,

criando artefatos que representam entidades reais. As principais características inerentes ao paradigma orientado a objetos que contribuíram para a nossa escolha são a *abstração* e a *herança*, que serão brevemente descritas a seguir.

- **Abstração** - É a característica que permite concentrarmos no que um objeto é e faz antes de se decidir como ele será implementado. Cada instância de uma classe possui um comportamento, que é um conjunto de mensagens às quais ela pode responder. Existe uma separação clara entre a *interface externa* de uma classe e sua *implementação interna*, sendo que esta última é mantida "escondida". Desta forma, implementações alternativas podem ser usadas para uma mesma classe sem alterar sua interface.
- **Herança** – É a característica que permite que uma classe herde o comportamento (operações, métodos etc) de outras classes existentes, possibilitando o compartilhamento do código e, desta forma, a reutilização. A herança é conseguida através da especialização das classes existentes, que pode acontecer pela extensão da representação ou do comportamento delas, ou ainda pela restrição de sua representação ou comportamento.

Para a documentação dos artefatos produzidos nas fases de análise e projeto, utilizamos os principais padrões da UML, como, por exemplo, o diagrama de classes para a modelagem conceitual, e o diagrama de casos de uso para a descrição das funcionalidades do sistema.

4.3.2 – Modelagem conceitual

Na modelagem de dados do ambiente CEJ foi utilizado o diagrama de classes da UML e a língua inglesa nas descrições textuais, com o intuito de garantir a internacionalização deste trabalho. Uma descrição detalhada desta modelagem é fornecida no apêndice B desta monografia.

4.3.3 – Uso do padrão de projeto MVC “Model 2”

Desde os primórdios do projeto do ambiente CEJ, estabelecemos como um objetivo importante torná-lo um sistema flexível e robusto, isto é, facilmente modificável. Para atingir este objetivo, decidimos pela escolha de um *framework* baseado na arquitetura *MVC* “*Model 2*”: o *framework Struts*. A aplicação do *Struts* à implementação do ambiente CEJ será descrita mais adiante neste capítulo, na seção 4.5.1.

Na literatura, um *framework* é descrito como uma técnica da Engenharia de Software para produzir uma arquitetura reutilizável e “semi-completa”, que pode ser especializada para construir aplicações customizadas (JOHNSON, 1998). O reuso suportado por um *framework* é de granularidade muito maior do que o oferecido por classes, contribuindo desta forma para um desenvolvimento mais organizado de sistemas de grande porte. No contexto mais específico dos sistemas orientados a objetos, FOLDOC (1995) define um *framework* como “um conjunto de classes que embute um projeto abstrato para a solução de problemas relacionados”.

O *framework Struts* baseia-se numa variação do padrão de projeto *MVC*. O modelo do *MVC* apresenta três componentes distintos: um componente para os objetos da aplicação (o *modelo*), um componente para a apresentação da interface do usuário (a *visão*), e um componente para o controle das requisições dos usuários (o *controlador*). Mas existem duas variações distintas para o padrão *MVC*, que são conhecidas como “*Model 1*” e “*Model 2*”.

Na arquitetura “*Model 1*”, um único componente é responsável pelo processamento das requisições do usuário e pela apresentação dos resultados. Mesmo assim existe uma separação da apresentação dos resultados ao usuário da lógica da aplicação, já que o acesso aos dados é feito de forma independente. Em um ambiente de desenvolvimento *Java*, tipicamente esse componente seria representado por uma única página *JSP*, enquanto num ambiente “*Microsoft*” ele seria representado por uma página *ASP*.

Embora esta arquitetura seja útil para aplicações simples, ela se torna desvantajosa quando se deseja dividir uma equipe em desenvolvedores e *designers* gráficos, responsáveis pela confecção das páginas *Web*. Neste caso, torna-se difícil a alocação de papéis e responsabilidades, já que o processamento da aplicação e a

apresentação ficam juntos num único local. A arquitetura “*Model 1*” é ilustrada na figura 4.1:

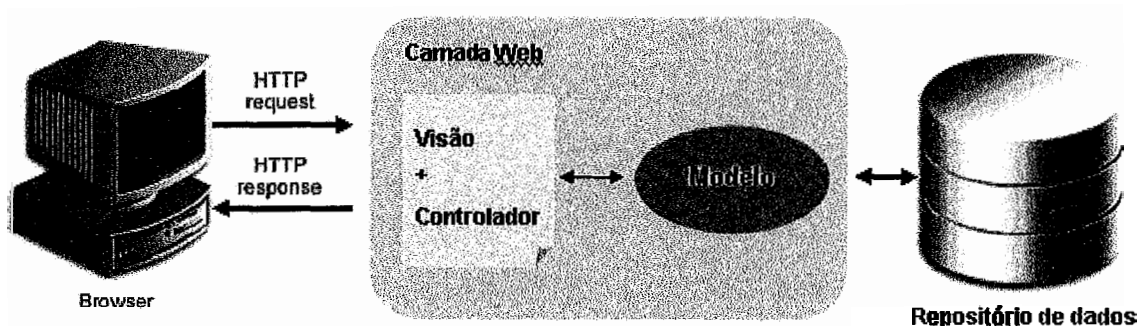


Figura 4.1 – Arquitetura MVC “Model 1”

Na arquitetura "*Model 2*", ilustrada na figura 4.2, a *visão* e o *controlador* são separados. O controlador, ao receber e processar uma requisição do usuário decide qual será a visão a ser retornada ao usuário, dependendo das ações do usuário. No componente de *visão*, não existe nenhum processamento lógico, ocorrendo apenas a apresentação dos objetos de exibição produzidos pelo controlador no atendimento da requisição do usuário.

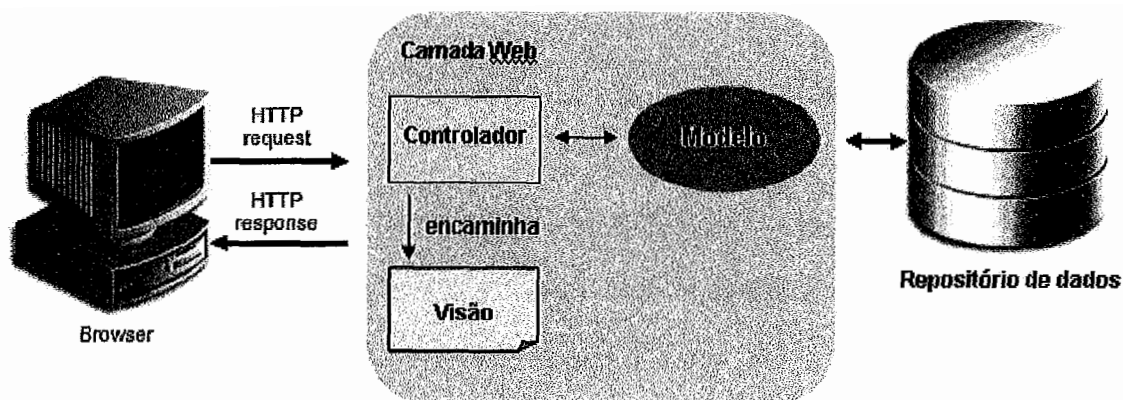


Figura 4.2 – Arquitetura MVC “Model 2”

Esta abordagem permite uma melhor definição dos papéis e responsabilidades dos desenvolvedores e *designers* de páginas *Web*, e quanto maior a complexidade da

aplicação, maiores os benefícios desta arquitetura. Uma outra vantagem deste modelo é melhor organizar sistemas que oferecem múltiplas apresentações de uma mesma informação, indo de encontro com os requisitos do ambiente CEJ.

4.3.4 – Arquitetura do ambiente CEJ

A arquitetura do ambiente CEJ é constituída de seis componentes básicos: um componente de *navegação personalizada e anotações colaborativas*, um componente para a *busca de documentos e anotações*, um terceiro componente para a *gerência do ciclo de vida dos documentos*, um quarto componente para a definição e execução de *workflows customizáveis de publicação*, um quinto componente de *notificações personalizadas*, e finalmente um componente para a *administração e configuração* do ambiente e das revistas que nele rodam. Estes componentes podem ser visualizados na figura 4.3 e serão descritos detalhadamente na seção seguinte.

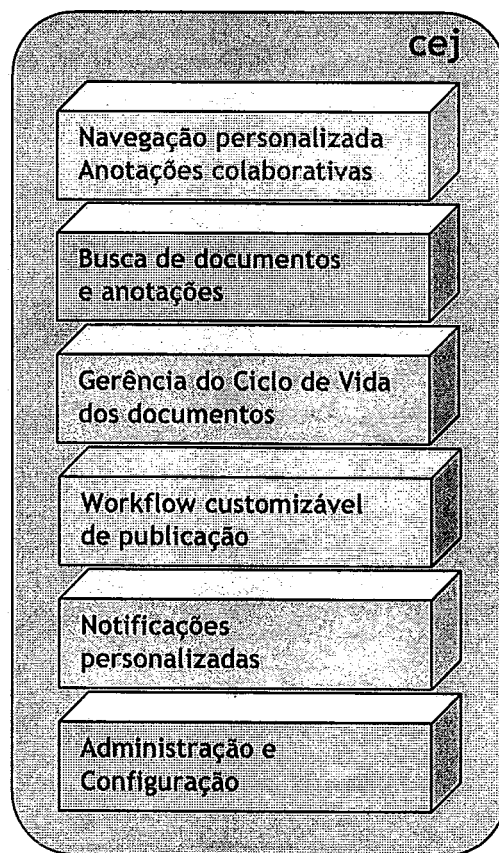


Figura 4.3 – Componentes da arquitetura do ambiente CEJ

4.4 – Descrição dos componentes da arquitetura do ambiente CEJ

4.4.1 – Componente de “Navegação e Anotação”

Os principais objetivos deste componente dentro da arquitetura do ambiente CEJ incluem: 1) a exibição dos documentos, no layout e formatação indicados pelo *template* escolhido pelo leitor; 2) o controle da navegação dos usuários sobre os documentos; 3) o suporte às anotações colaborativas.

4.4.1.1 – Navegando sobre um documento

Há várias maneiras de explorar um documento submetido ao ambiente CEJ. Um documento pode ser exibido nos formatos XHTML (formato padrão), XML (formato nativo) e PDF (formato para impressão), conforme sugerido na seção 4.2.5. A figura 4.4 ilustra os modos de navegação sobre um documento no ambiente CEJ:

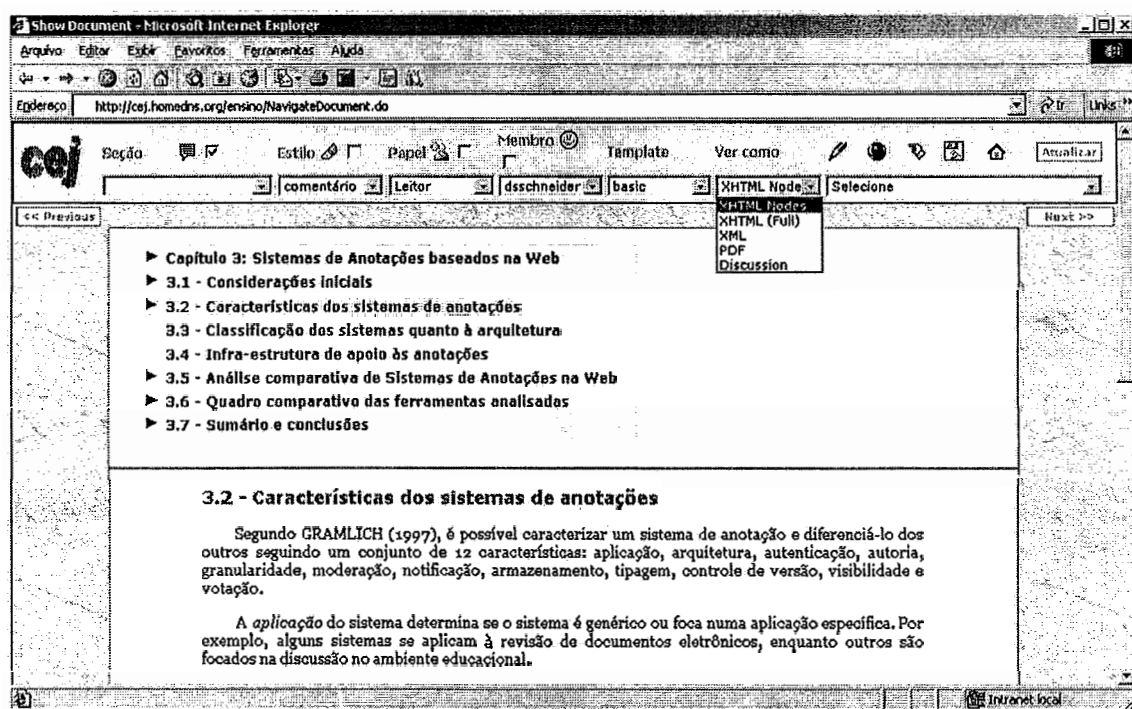


Figura 4.4 – Modos de navegação sobre um documento

A exibição padrão de um documento ocorre no formato XHTML, que permite que ele seja incrementado com anotações e comentários. Neste caso, o sistema realiza um *merge* do documento original com um *template* de exibição escolhido pelo leitor. Há dois modos de visualizar um documento no formato XHTML: 1) como um único objeto; 2) como diversos nós interligados. É possível também visualizar o documento com ou sem as anotações, conforme será discutido na seção 4.4.1.3.

Com a definição e utilização dos *templates* de exibição customizáveis é possível, por exemplo, visualizar somente o resumo dos artigos - como já acontece com muitas revistas eletrônicas - ou navegar somente pelas figuras de um documento. É possível também associar os nós de um artigo a determinados eventos, trazendo maior interatividade à navegação sobre o artigo. Por exemplo, poder-se-ia associar a figura de uma fórmula matemática a um evento de clique, que quando acionado exibiria uma demonstração da fórmula ou permitira o teste da função para um conjunto de parâmetros informados pelo leitor.

Outro modo de visualização apresenta somente as referências do documento. O sistema exibe não apenas a lista dos documentos referenciados pelo presente documento, como a lista dos documentos que referenciam tal documento. Esta lista é um *merge* da lista de documentos armazenados lógicamente ou fisicamente no servidor do CEJ, com a lista de documentos encontrada na *Web*, através de motores de busca como o *Google*. Desta forma, este componente atende ao requisito de suportar *referências bidirecionais*, conforme proposto anteriormente. Apontadores para as referências são criados automaticamente por este componente, e poderão encaminhar o usuário ao próprio ambiente CEJ – caso o sistema identifique o documento – ou para outros servidores.

Quando o documento é visualizado na forma de *nós interligados*, o sistema constrói automaticamente um sumário com os elementos do primeiro nível do documento, que correspondem tipicamente aos metadados do documento, às seções e à bibliografia, quando existir. Além disso, à medida que um leitor vai explorando cada um dos nós do documento, o sistema calcula e informa ao leitor o percentual do documento que já foi lido. Opcionalmente, o leitor poderá marcar que nós ele já explorou, e essa informação poderá ser enxergada por outros membros da revista.

A granularidade mais fina possível para uma anotação corresponde à de uma letra selecionada de um trecho qualquer do documento. Em seguida, para adicionar o título e conteúdo da anotação, o usuário deverá clicar em cima do trecho selecionado, já “marcado”, e o sistema abrirá um *frame* – embutido no próprio documento – para que sejam informados estes dados, conforme ilustra a figura 4.5. Finalmente, deve-se clicar o botão “Adicionar” para que a anotação seja criada no repositório do sistema.

Para navegar sobre uma anotação, o usuário deverá clicar em cima da âncora da anotação escolhida. Neste caso, o sistema também abre um *frame* dentro do próprio documento, nas proximidades da porção do documento anotado. Neste *frame* são exibidos os atributos da anotação – como o título, o corpo e o autor – e uma lista das respostas a esta anotação. Para navegar sobre alguma destas respostas, o usuário deverá selecionar a resposta desta lista e clicar o botão “Ok”. Sobre cada uma das respostas da anotação original também poderão ser criadas anotações, de forma que é possível criar uma árvore ou linha de discussão com um nível arbitrário de níveis. A figura 4.6 ilustra a navegação de um usuário sobre uma anotação localizada.

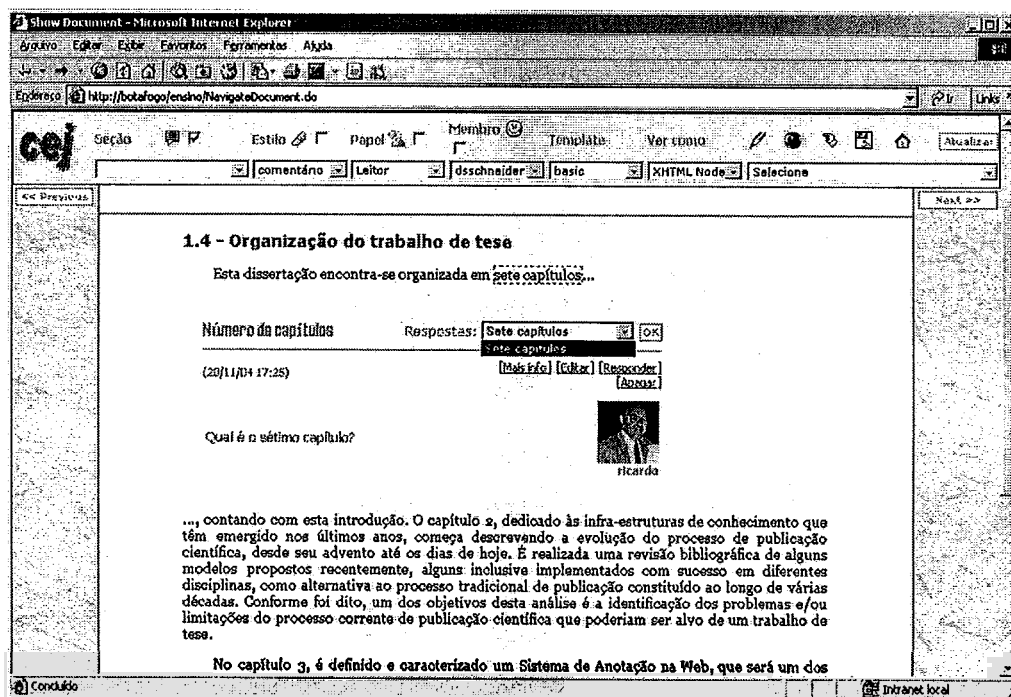


Figura 4.6 – Navegando sobre uma anotação localizada no ambiente CEJ

Observe que existe uma resposta a esta anotação, no canto superior à direita do *frame* embutido no documento.

Comentários gerais sobre o documento como um todo também podem ser adicionados, mas para isso o usuário deverá visualizar o documento no modo “Discussão”. Neste caso, âncoras para as anotações não fazem sentido, e o sistema exibe todos os comentários no espaço à direita do documento.

4.4.1.3 – Filtrando anotações

Considerando um documento com um grande número de anotações localizadas, pode ser inconveniente visualizar todas as anotações simultaneamente. Por outro lado, um determinado membro leitor de uma revista poderá estar interessado em analisar somente um subconjunto restrito das anotações – por exemplo, as anotações feitas pelos “revisores” da revista.

Desta forma, o componente de navegação e anotações do ambiente CEJ oferece mecanismos de filtro, que permitem esconder do usuário as anotações não desejadas, de acordo com os seguintes critérios: tipo de anotação, papel do usuário criador da anotação e nome do usuário criador, o que é claramente visível na interface de navegação sobre os documentos (figura 4.7). Opcionalmente, o usuário poderá esconder todas as anotações, visualizando somente o documento original.



Figura 4.7 – Filtros de visualização das anotações no ambiente CEJ

A figura acima ilustra o painel superior da interface de navegação sobre documentos do ambiente CEJ. Observe que no exemplo o usuário está filtrando as anotações pelo estilo “iluminador” e pelo membro da revista “ricardo”, criador da anotação.

4.4.2 – Componente de “Busca de documentos e anotações”

Este componente oferece um serviço de busca de documentos e anotações armazenadas (lógica ou fisicamente) no ambiente CEJ. Um diferencial deste serviço de

busca está na possibilidade de buscar palavras ou expressões em um contexto específico do documento - que é possível em função de os documentos serem submetidos no formato XML e armazenados de forma estruturada e indexada no repositório de dados.

Por exemplo, o usuário poderá buscar a expressão “ontologias de domínio” nos parágrafos dos documentos, dentro do resumo (*abstract*), ou ainda no título do documento. Uma busca por autor ou por título também seria muito precisa do que uma busca realizada em qualquer motor de busca disponível na *Web*, onde os elementos de um documento (seções, parágrafos etc) não recebem uma formatação especial descrevendo sua semântica.

Outro serviço oferecido pelo componente é a busca por anotações baseada em seus atributos, como autor da anotação, título, corpo ou contexto. Este poderá ser um recurso valioso ao evitar o trabalho desnecessário dos pesquisadores na leitura e navegação sobre documentos que não se encaixem no seu perfil ou interesse.

A versão atual deste componente ainda é incipiente, suportando somente uma parte das funcionalidades idealizadas, mas deverá ser aprimorada no futuro próximo.

4.4.3 – Componente de “Gerência do ciclo de vida”

O principal objetivo deste componente dentro da arquitetura do ambiente CEJ, como o próprio nome também indica, é controlar o ciclo de vida dos documentos submetidos a uma determinada revista. O ciclo de vida de um documento neste ambiente poderá envolver diversas etapas, como *criação*, *submissão*, *categorização*, *publicação* e *eliminação*.

Este ciclo inicia-se com a **criação** do documento. Esta etapa poderá ser realizada dentro ou fora do sistema, mas o ambiente CEJ suportará a criação de documentos no formato HTML e sua conversão automática para o formato XML, conforme comentado anteriormente neste capítulo. Esta conversão é ilustrada na figura a seguir, que apresenta o “Gerador de XML” do ambiente CEJ:

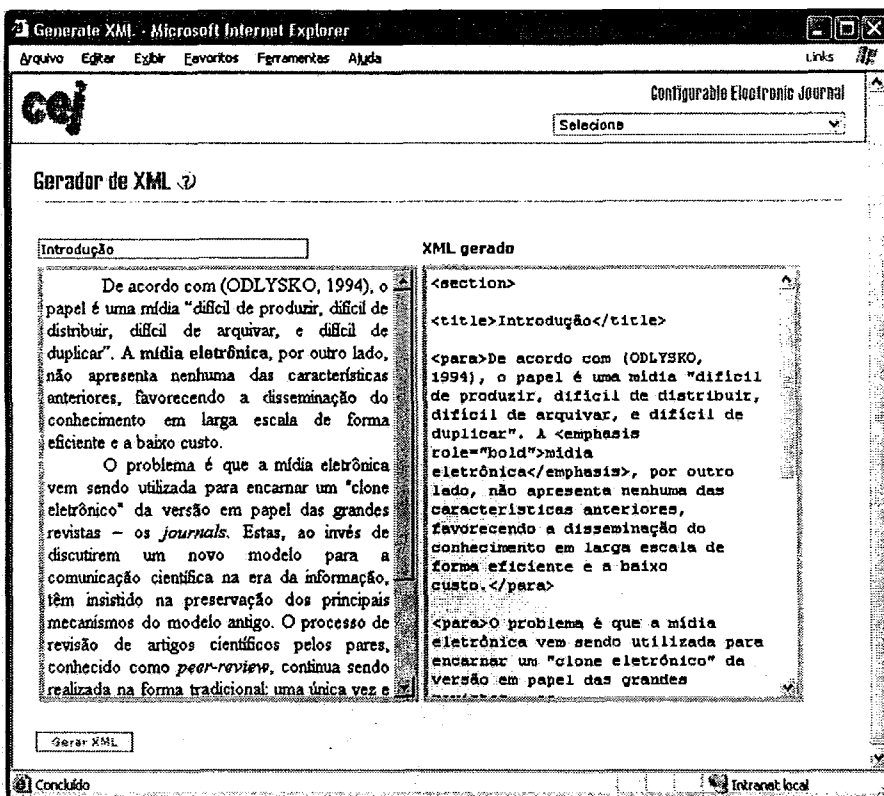


Figura 4.8 – Conversor de HTML para XML do ambiente CEJ

A **submissão** dos documentos possibilita sua inclusão física no ambiente CEJ, ou ainda a inclusão de um ponteiro (*URL*) para um dado documento. Isto acontecerá tipicamente através da execução de uma tarefa do tipo “submissão”, no contexto de um processo de publicação. A submissão não significa publicação, isto é, não implica em o documento estar disponível para leitura na revista, a não ser que a submissão seja sucedida imediatamente por uma ação de publicar o documento.

No projeto do ambiente CEJ, como comentamos anteriormente, foi escolhido o padrão *XML* para a submissão de documentos, por possibilitar a separação do conteúdo de um documento da sua apresentação. No primeiro protótipo do CEJ, decidimos adotar o *DocBook*, um formato *XML* de código aberto para a publicação de livros e artigos. De maneira semelhante ao que acontece com a revista *PubMed Central*, descrita no capítulo 2 desta monografia, este componente do ambiente CEJ oferece uma API para a validação dos documentos *XML*, com o objetivo de realizar verificações a respeito da completude e corretude sintática dos documentos. Esta API poderá ser utilizada, por

exemplo, para garantir que todo artigo em uma revista científica contém pelo menos um autor, que o resumo do artigo é obrigatório, e que o corpo do documento (suas seções) não ultrapasse o tamanho de 5000 palavras. Além disso, uma página com as “Instruções para os autores” poderá ser gerada automaticamente pelo sistema apresentando estas regras definidas pelo editor da revista. Finalmente, um documento submetido no formato XML é armazenado de forma estruturada em um repositório de dados, gerenciado por um SGBD relacional.

Em seguida à submissão será feita uma **categorização** do documento, com o cadastro de um conjunto de metadados deste documento. Uma parte dessa categorização será feita automaticamente, já que muitas informações poderão ser obtidas do próprio documento, como nome e endereço dos autores, categoria do artigo etc. Esta categorização é *interativa*, isto é, o sistema deduz as informações pertinentes do documento XML submetido, mas permite aos autores (ou ao editor) a edição dessas informações para fins de catalogação.

Conforme discutido anteriormente, a **publicação** de um documento no ambiente CEJ a princípio poderá ocorrer em qualquer instante, e mesmo antes da revisão, situação que se convencionou chamar de “publicação instantânea”. Deste modo, o componente de gerência do ciclo de vida deverá associar um documento a determinadas *estampas*, que indicarão ao usuário se o documento já foi revisado, já foi aprovado pelo editor da revista etc. Isto é feito automaticamente, facilitando o trabalho dos editores da revista. Um histórico também é feito das mudanças de *status* de um documento de acordo com seu ciclo de vida, e os membros da revista poderão ser informados dos eventos de mudança.

Finalmente, poderá acontecer a **eliminação** de um documento, a última etapa que ele pode passar no ciclo de vida. Vimos no capítulo 2 e na seção 4.2.1 deste capítulo que a publicação de artigos deve ser *reversível* no ambiente eletrônico. Desta forma, é possível remover um documento por completo da revista, caso a revisão do mesmo não tenha sucesso. Como vimos, isso pode acontecer em função da rejeição imediata do artigo pela equipe de revisão, por causa do não-atendimento dos pedidos de modificação por parte dos autores do documento, ou em função da própria desistência dos autores em continuar o ciclo de revisão.

Vale aqui ressaltar que no momento de eliminação de um documento, o editor poderá optar pela exclusão lógica ou física do documento. No primeiro caso, o documento – ou o ponteiro para a URL do mesmo - fisicamente permanecerá armazenado por um certo período de tempo. Posteriormente, o editor deverá acessar a interface de administração e configuração do ambiente para removê-lo definitivamente.

Por último, é conveniente enfatizar o aspecto de suporte ao *versionamento* de documentos no ambiente CEJ. Tal suporte, considerado dentro do contexto de uma revista de publicação eletrônica, está intimamente associado à capacidade de alteração de um documento depois de publicado. Como vimos no capítulo 2, a maioria das revistas não suporta a edição de documentos após a publicação, mas normalmente é possível revisar um dado documento um número arbitrário de vezes. Deste modo, no ambiente CEJ, o suporte ao versionamento se limita à geração de novas versões do documento. Cada nova versão do documento poderá receber anotações, mas as anotações localizadas de uma versão anterior não farão sentido nas outras versões.

A investigação da conveniência de permitir a alteração dos documentos pelos próprios autores, associada ao versionamento dos mesmos, e à capacidade de transferir as anotações de uma versão para outra do documento, é um tarefa laboriosa que será alvo de um trabalho futuro.

4.4.4 – Componente de “Gerenciamento de Workflows”

O quarto componente da arquitetura do CEJ é um gerenciador de *workflows* customizáveis de publicação. Este gerenciador inclui uma API tanto para a definição quanto para a execução dos processos de publicação.

- Definição de processos de publicação

Atualmente a definição de processos de publicação no ambiente CEJ acontece de forma não usual, já que os processos, as tarefas, os *links*, os participantes e demais objetos são definidos através de uma interface exclusivamente textual, ilustrada na figura 4.9. A construção de uma interface gráfica para a criação destes objetos será alvo de um trabalho futuro.

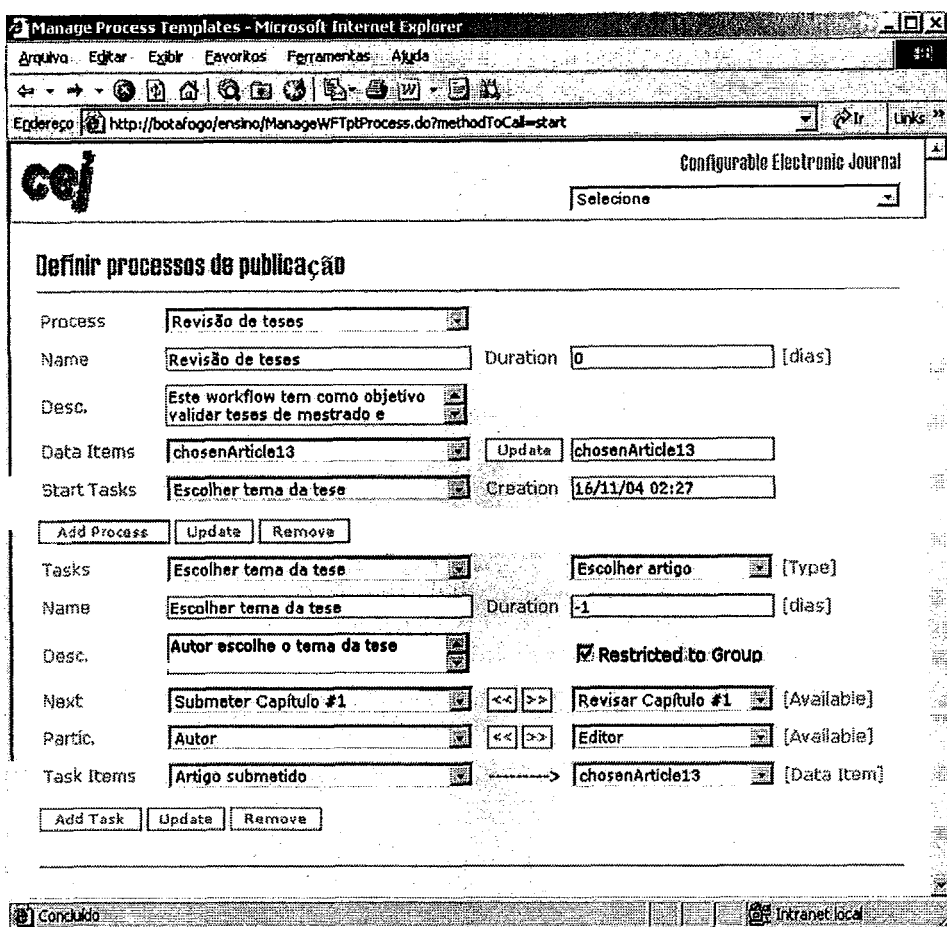


Figura 4.9 – Definição de processos de publicação no ambiente CEJ

É importante destacar que a definição das tarefas de um *template* de processo de publicação é condicionada a determinadas restrições ou pré-requisitos, associadas aos parâmetros de entrada e saída do tipo de tarefa. Por exemplo, a primeira tarefa de um *template* de processo não poderá jamais ser do tipo "Publicar documento", pois este tipo de tarefa exige como parâmetro de entrada um documento a ser publicado. Desta forma, esta tarefa deverá necessariamente ser antecedida por alguma tarefa cujo tipo defina como um dos seus parâmetros de saída um documento, por exemplo, uma tarefa do tipo "Submeter documento".

Parece evidente que a própria interface de configuração de *templates* de processos de publicação deve gerenciar estas restrições, somente permitindo a escolha de tipos de tarefa que atendam aos pré-requisitos exigidos.

- Execução de processos de publicação

No ambiente CEJ, a execução de processos de publicação ocorre através do caso de uso “Participar de processos de publicação”. Esta interface permite a instanciação dos *templates* de processos de publicação em *casos de processos*, e controla também a execução das tarefas que compõem o caso.

Uma síntese da documentação da API do componente gerenciador de workflows do ambiente CEJ, incluindo suas classes e a interface dos atributos e métodos de cada classe, poderá ser encontrada no portal do ambiente CEJ. A figura a seguir ilustra a interface para a execução dos processos de publicação através do componente de *workflows*.

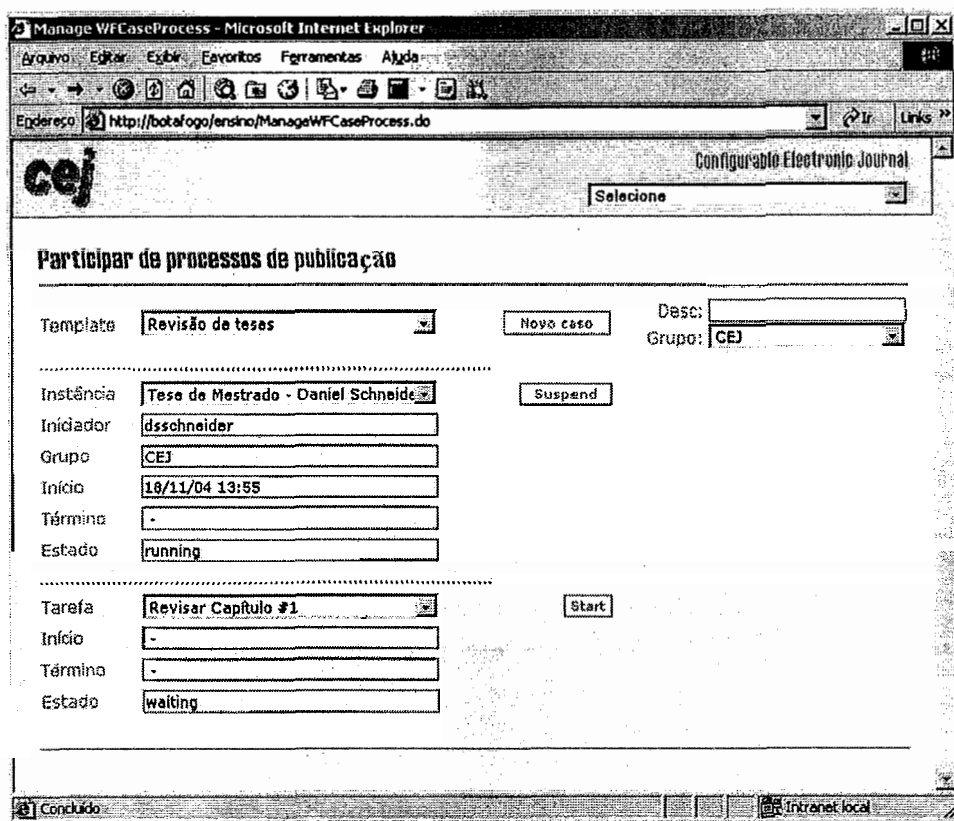


Figura 4.10 – Execução de processos de publicação no ambiente CEJ

4.4.5 – Componente de “Notificações personalizadas”

O quinto componente da arquitetura do CEJ tem como principal objetivo gerenciar a configuração dos eventos de notificação de uma determinada revista, e possibilitar o envio destas mensagens. Estas notificações podem ocorrer através do correio eletrônico, ou de uma lista personalizada ao membro da revista, exibida em uma página específica do sistema na *Web*. Estes eventos de notificação incluem, mas não se limitam aos seguintes:

- Envio de mensagem para o usuário que criou uma anotação sobre um documento.
- Envio de mensagem para o usuário que teve alguma anotação sua respondida.
- Envio de mensagem para os membros da revista, quando surgem novos documentos no repositório, ou quando muda o *status* de um documento – por exemplo, quando uma revisão é concluída.

Atualmente este componente só oferece suporte às notificações através do correio eletrônico, mas a extensão de suas funcionalidades deverá acontecer no futuro próximo. Um outro trabalho futuro é investigar formas mais flexíveis de definição dos eventos de notificação, dado que no estágio atual da implementação é exigido um esforço do desenvolvedor para incluir novos eventos.

4.4.6 – Componente de “Administração e Configuração”

O último componente da arquitetura do CEJ tem como objetivos a administração e a configuração dos principais objetos que participam do ambiente, como revistas, usuários, documentos, *templates* de exibição de documentos etc.

Um fato que deve ser destacado é que este componente não limita o uso das funcionalidades de configuração oferecidas aos editores. Por exemplo, na configuração de um *tipo de anotação*, o sistema permite que membros da revista (habilitados) adicionem, modifiquem e excluam um tipo. A exclusão e a modificação de um tipo de anotação seguem a filosofia “quem criou é quem pode excluir/modificar”. Para evitar que a remoção de um tipo de anotação removesse em cascata todas as anotações feitas com este tipo, adotamos a política “*on delete set default*”, ou seja, quando um tipo de

anotação é removido, as anotações que apontavam para aquele tipo passam a apontar para um tipo *default*, definido pelo editor da revista. A mesma política vale para a configuração dos *templates* de exibição dos documentos do ambiente CEJ. A figura a seguir ilustra a configuração dos tipos de anotação e dos *templates* de exibição no ambiente CEJ:

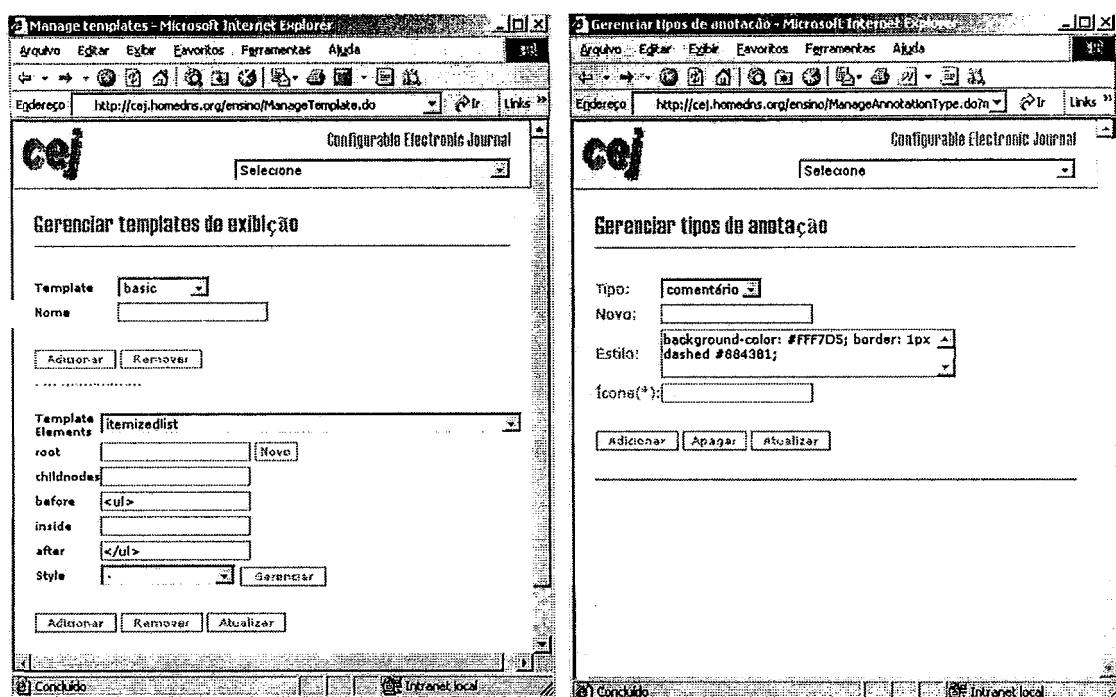


Figura 4.11 – Configuração de *templates* de exibição e de tipos de anotação no CEJ

Um outro objetivo não menos importante deste componente é o de fornecer mecanismos para o gerenciamento das funcionalidades oferecidas pelo sistema - os casos de uso - e o controle do acesso a tais funcionalidades. Tal componente suporta a inclusão, a adição e a remoção inteligente dos casos de uso do sistema, já que um *caso de uso* é um objeto como qualquer outro do ambiente CEJ. Neste caso, o acréscimo de novas funcionalidades e o controle de acesso a tais funcionalidades – os casos de uso – é simplificado, já que o sistema oferece suporte completo à sua configuração pelo ambiente *Web*.

A figura a seguir ilustra este suporte do ambiente CEJ à configuração dos casos de uso. Todavia, a criação dos artefatos necessários à implementação ou alteração de um

caso de uso deverá ser feita através de outra ferramenta – o “portal de desenvolvimento do ambiente CEJ” - o que será discutido mais adiante na seção 4.5.4.

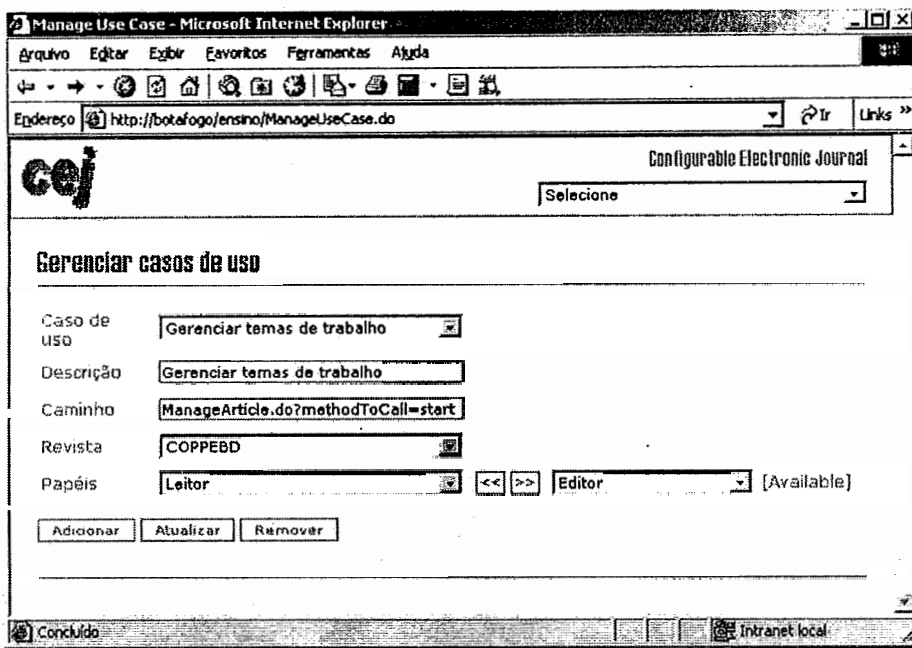


Figura 4.12 – Configuração dos casos de uso do ambiente CEJ

4.5 – Implementação e protótipo

Nesta seção serão descritas as principais abordagens e ferramentas escolhidas para a implementação do ambiente CEJ. A primeira versão experimental deste ambiente, na qual foram realizados os experimentos que serão descritos no capítulo seguinte, está disponível na *Web* em <http://cej.homedns.org/ensino>, e pode ser acessada por qualquer leitor anônimo utilizando o nome de usuário “*guest*” e a senha “*convidado*”. Ali estão disponíveis os artigos que foram publicados e em todos os experimentos realizados até aqui com este ambiente.

4.5.1 – Java e o framework *Struts*

O *framework Struts* é uma tecnologia aberta para o projeto e a construção de aplicações *Web* escritas na linguagem Java. Alguns pontos justificaram a adoção deste pacote na implementação do ambiente CEJ:

- É uma tecnologia gratuita, de código aberto e bastante documentada através de livros, manuais e artigos disponíveis na *Web*.
- Todos os componentes do Struts podem ser estendidos - através do mecanismo de herança básico da linguagem Java - e substituídos.
- Inclui poderosas bibliotecas de tags que podem ser utilizadas para a geração de visões em JSP onde não existe nenhum código Java. As mais conhecidas são as bibliotecas *Struts-Html*, a *Struts-Bean* e a *Struts-Logic*.
- Oferece suporte a diversas ferramentas e tecnologias de outros fornecedores que podem ser acopladas ao framework: a ferramenta IBM WebSphere Studio Advance Developer (WSAD), Enterprise Developer (WSED), Eclipse, Jbuilder etc.

O Struts roda em qualquer plataforma (*Windows, Linux, Solaris* etc), e pode ser acoplado em diversos servidores de aplicação, como o *Tomcat*, o *Resin* e o *WebLogic*. Numa primeira etapa do desenvolvimento do ambiente CEJ, testamos o *Struts* com o *Apache Tomcat*. Numa segunda etapa optamos por trocar o *Tomcat* pelo servidor de aplicação *Resin*, que oferece algumas vantagens em comparação com o *Tomcat*, das quais a principal é a compilação automática e inteligente das classes Java de uma aplicação.

A figura 4.13 apresenta a arquitetura MVC “Model 2” de implementação do ambiente CEJ, de acordo com o framework Struts. Note que ambiente baseia-se numa solução cliente-servidor *Web*, e que não é necessário nenhum outro tipo de software especial no lado cliente além de um navegador *Web* padrão, como o *Netscape* ou o *Internet Explorer*. Uma requisição do cliente é sempre interceptada por um componente do tipo *servlet*, conhecido como *ActionServlet* (o controlador). Este componente processa esta requisição e determina qual a visão que deve ser encaminhada ao usuário. Note ainda que, no *framework Struts*, o cliente nunca envia uma requisição diretamente a uma visão. Isto permite que o controlador centralize em um único local toda a lógica de controle da aplicação. Uma vez que a requisição tenha sido processada pelo controlador, este direciona a requisição à página apropriada (a visão).

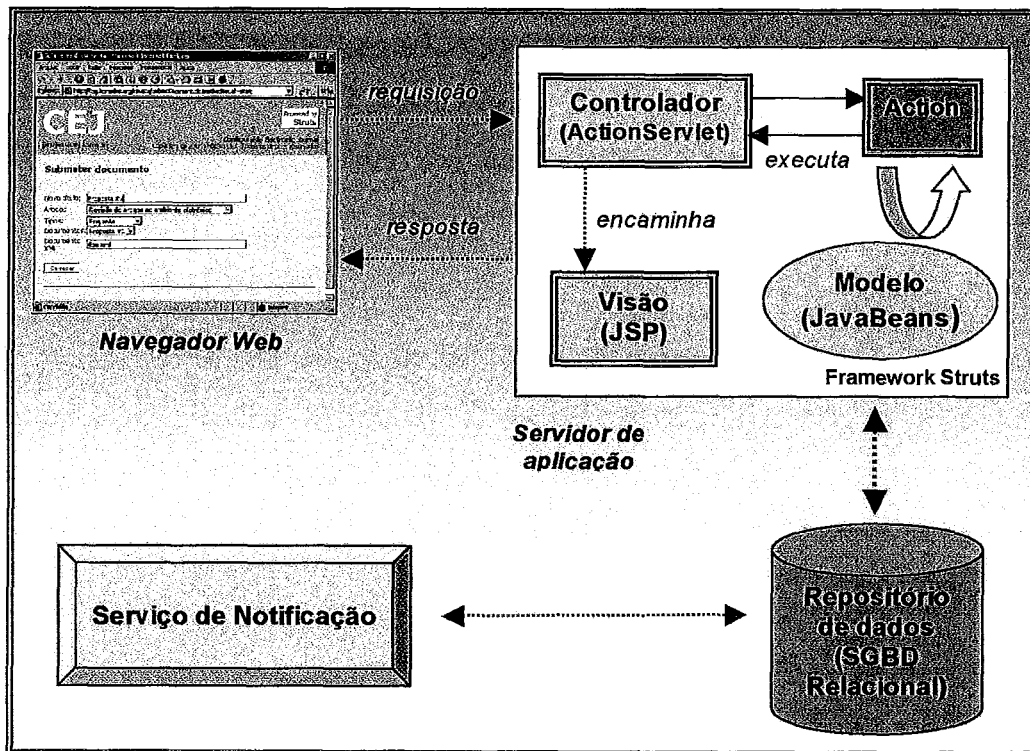


Figura 4.13 – Arquitetura MVC “Model 2” de implementação do ambiente CEJ

O ambiente CEJ foi construído com classes Java reutilizáveis e desenvolvidas independentemente, o que justifica o potencial de customização, modificação e extensão das funcionalidades deste sistema. Conforme se pode notar pela figura 4.13, as classes do componente “modelo” interagem diretamente com o repositório de dados, de tal forma que o modelo é construído também a partir de classes Java. Uma outra possibilidade, útil para aplicações mais complexas, seria a de implementar o modelo através da tecnologia de *Enterprise JavaBeans (EJB)*, onde o componente do modelo comunicaria-se com um servidor de *EJB*. Esta abordagem, todavia, foi descartada por simplicidade. Para o repositório de dados, escolhemos o SGBD relacional *MySQL*, mas qualquer outro sistema gerenciador de banco de dados poderia ser acoplado (*Oracle*, *PostgreSQL* etc), contanto que oferecesse suporte à tecnologia *JDBC*.

As visões dentro da arquitetura do *Struts* são representadas por páginas *JSP*, incrementadas com os *tags* do *Struts*, que podem ser utilizadas tanto na exibição de conteúdos estáticos quanto dinâmicos. Praticamente todo o conteúdo dinâmico que é exibido em uma visão é produzido no lado servidor, pela aplicação *Web*. Todavia, uma

pequena quantidade de código no ambiente CEJ é escrita na linguagem *JavaScript* do lado cliente, especialmente na implementação do componente de navegação e anotações colaborativas. Finalmente, há ainda um serviço de notificação na arquitetura do CEJ, também implementado em Java, que pode ser configurado para gerar automaticamente mensagens, através do correio eletrônico, para os membros da revista, conforme discutido na seção 4.4.5.

Na construção do protótipo do ambiente CEJ, o lado cliente da implementação do componente de navegação e anotações foi construído utilizando a API de acesso ao modelo de objetos do documento (*DOM*) do navegador *Microsoft Internet Explorer*. Esta escolha se justifica pela flexibilidade oferecida por esta API, sobretudo na geração das âncoras das anotações localizadas, e ao fato de ser a API do navegador “oficial” do ambiente *Windows*, o mais popular atualmente entre os usuários do globo. Deste modo, o suporte completo às funcionalidades do ambiente CEJ limita-se a este navegador, na versão experimental deste ambiente. Como planos futuros, descritos no capítulo 6, está a compatibilização da API do CEJ com outros navegadores, como a família de navegadores baseados no *Mozilla*, incluindo o *Netscape*.

4.5.2 – Implementação do “Gerenciador de Workflows”

O componente gerenciador de workflows, que permite tanto o controle da definição quanto da execução de processos de publicação, foi projetado e implementado com as seguintes características:

1) **Orientado a objetos** – Isto significa que o componente gerenciador de *workflows* também segue o paradigma principal de construção do ambiente CEJ: a orientação a objetos. Deste modo, *templates* de processos, de tarefas e de *links*, assim como suas respectivas instâncias, são tratados como objetos com atributos, métodos e interfaces.

A figura 4.14 ilustra o esforço requerido para desenvolver uma solução de *workflow* contrastado com a flexibilidade da solução. O aumento da flexibilidade implica o aumento da complexidade. Se o sistema é muito flexível e requer muita programação, será difícil também de manter. Por outro lado, se as primitivas em um sistema são muito restritivas, será difícil implementar uma solução que resolva os

problemas que se deseja automatizar (KHOSHAFIAN et al., 1995). As soluções que envolvem maior esforço de programação, de acordo com a figura, são as construídas “do zero” em uma arquitetura cliente-servidor. Com a orientação a objetos ocorre um incremento da flexibilidade, mas o esforço de programação ainda é requerido nos níveis mais baixos. De acordo com KHOSHAFIAN et. al. (1995), a *orientação a objetos* combinada com *regras declarativas*, geralmente fáceis de implementar, oferecem a melhor solução em termos de flexibilidade e esforço de desenvolvimento.

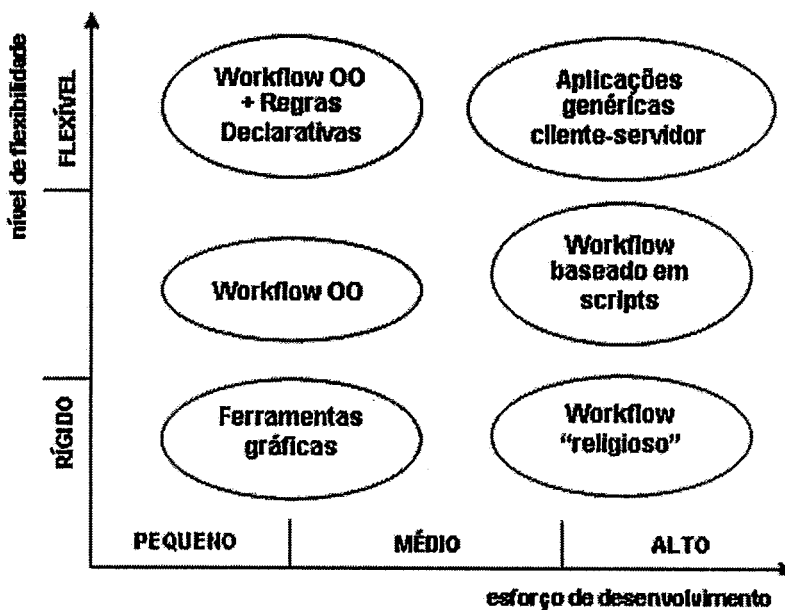


Figura 4.14 – Esforço x flexibilidade na solução de um *workflow* (KHOSHAFIAN, 1995)

Na implementação do gerenciador de *workflows* do ambiente CEJ, optamos somente pelo uso da orientação a objetos, em concordância com o *framework* Struts. Um dos casos de uso do sistema, acessível somente a usuários com papel de "desenvolvedor", permite que sejam criados novos "tipos de tarefa", permitindo aumentar o conjunto básico de atividades que poderão compor um *workflow* de publicação. Um novo tipo de tarefa poderá ser definido de duas formas: estendendo qualquer um dos tipos de tarefa pré-definidos, ou criando um novo tipo de tarefa "do zero". Para implementar com sucesso um novo tipo de tarefa, tudo que é necessário é fornecer os ingredientes exigidos pela interface de inclusão de um novo tipo, como o

nome do tipo da tarefa e o caminho - de acordo com o framework Struts - e escrever uma classe Java que implemente este novo tipo - através do portal de desenvolvimento do ambiente CEJ.

Tudo isto pode ser feito pelo próprio navegador Web, através do portal da revista e do portal de desenvolvimento. O sistema reconhece automaticamente o novo tipo de tarefa criado, já que a nova classe Java criada é compilada e carregada automaticamente pelo servidor de aplicações!

Urge destacar que qualquer novo tipo de tarefa deve ser criado herdando de um tipo de tarefa básico, o que permite reutilizar grande parte da complexidade na escrita de um novo tipo. Desta forma, não há razão plausível para "começar do zero", a não ser que o desenvolvedor queira construir algo totalmente novo e desviar-se do padrão de projeto planejado para as classes do gerente de *workflows*, o que não é recomendado.

2) **Baseado na Web** – Isto significa que a execução das atividades de um processo de publicação ocorre através de um navegador *Web*. Deste modo, não deve ser necessária nenhuma outra ferramenta para definir ou executar os processos de publicação, sendo a única exceção a atividade de composição dos documentos, que poderá ser feita parcialmente em um editor de texto.

Há seis anos atrás, CHAFFEY (1998) apresentava uma visão crítica e pessimista das ferramentas de workflow baseadas na Web, em comparação com outras ferramentas de *workflow* cliente-servidor não baseadas na Internet ou Intranet:

“Utilize qualquer uma das aplicações de demonstração (de workflows) acessíveis na Internet e você concluirá que a performance é pobre ou torna simplesmente a aplicação não-utilizável. Cada nova interação do usuário exige a carga de um novo formulário. Até mesmo a exibição de uma lista de opções simples pode levar segundos, enquanto numa aplicação cliente-servidor tradicional ela seria instantânea. Os atrasos encontrados tornam o uso de um workflow baseado na Web impraticável para usuários de um workflow de produção. Será necessário esperar pelas futuras gerações da Internet...” (CHAFFEY, 1998)

Todavia, a popularização da banda larga e os recentes avanços de hardware, como processadores mais rápidos, *clusters* de servidores e memórias cada vez mais abundantes, aliados aos mecanismos de gerenciamento de *threads* no lado servidor, que permitem a execução intercalada de tarefas, tornaram recentemente os *workflows* baseados na Web uma realidade. Os primeiros experimentos realizados no ambiente CEJ, que serão descritos no capítulo seguinte, sugerem que o *workflow baseado na Web* é uma solução plausível e ótima para o gerenciamento e a execução de processos de publicação científica.

4.5.3 – Implementação da API de geração de PDF

A implementação da API de suporte à geração da versão impressa dos documentos do ambiente CEJ merece também algumas palavras. A implementação de tal API requer a recuperação dos elementos do documento (originalmente em XML) e dos atributos dos elementos e a geração de um novo documento satisfazendo o padrão PDF. Isto a princípio poderia ser uma tarefa laboriosa, sobretudo porque o padrão interno do formato PDF não é documentado extensivamente. Todavia, acoplamos à nossa implementação uma biblioteca gratuita que possibilita a geração *on-the-fly* de documentos no formato PDF: a biblioteca iText (ITEXT, 2004). Esta é apenas uma das muitas razões que demonstram que a escolha da linguagem Java é uma decisão acertada, em função da grande quantidade de bibliotecas de componentes que estão disponíveis de forma gratuita na *Web*.

4.5.4 – Portal de desenvolvimento do ambiente CEJ

O “portal de desenvolvimento do ambiente CEJ” representa uma característica particular e inovadora do processo de construção da implementação deste ambiente. Trata-se de um ambiente de auxílio à implementação do código do CEJ, que da mesma forma que o ambiente CEJ foi implementado na plataforma cliente-servidor Web, e é acessado através de um navegador padrão. Este portal permite o desenvolvimento e a extensão do ferramental do CEJ, e inclui interfaces para a navegação sobre os artefatos de desenvolvimento, acesso ao SGBD incorporado ao ambiente, e liberação dos artefatos novos ou modificados para o ambiente de “produção”.

O acesso ao portal de desenvolvimento do CEJ é restrito a usuários cadastrados por um administrador, e diversas instâncias do ambiente CEJ podem ser acessadas através de um único portal. A figura a seguir ilustra a página inicial do portal, onde são exigidos o nome da instância da aplicação CEJ e a senha para acessar tal instância:

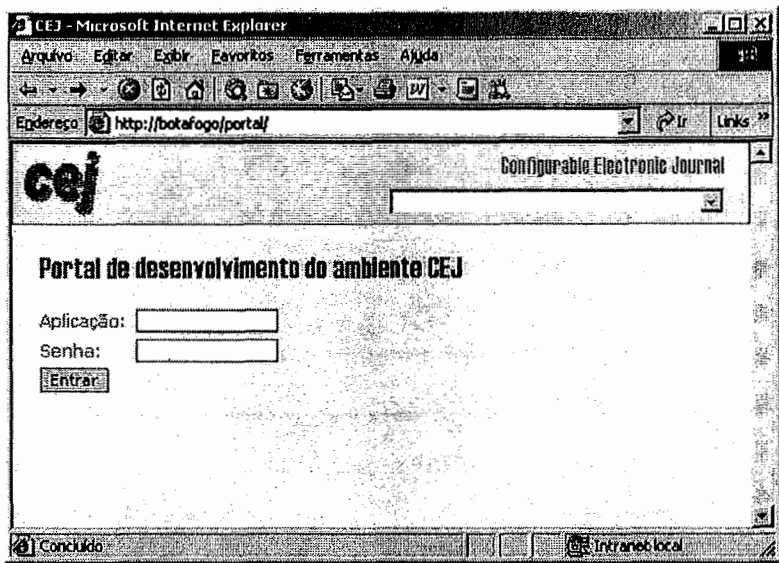


Figura 4.15 – Portal de desenvolvimento do ambiente CEJ

O portal de desenvolvimento do CEJ pode ser acessado através do apontador <http://cej.homedns.org/portal>. As duas principais interfaces deste ambiente serão descritas a seguir.

4.5.4.1 Navegação sobre os artefatos de desenvolvimento e compilação automática das classes Java

Esta interface permite navegar sobre a árvore de diretórios da aplicação Web correspondente à instância do ambiente CEJ escolhida. Todos os artefatos de desenvolvimento se encontram ali localizados, incluindo-se as visões em JSP, as classes Java de implementação, os documentos XML de configuração da aplicação e do controlador, e finalmente os objetos acessórios, como imagens, arquivos de estilos etc.

Esta interface permite não apenas a navegação sobre estes artefatos, como possibilita a adição, a remoção e a edição dos mesmos. Desta forma, ela foi empregada para o desenvolvimento do próprio ambiente CEJ pelo autor da monografia. Um aspecto que merece ser enfatizado é que a compilação das classes Java e das visões JSP, além da

carga dos documentos XML de configuração do controlador, são feitos automaticamente e de forma inteligente, graças ao uso do servidor de aplicações *Resin*. Por exemplo, toda vez que uma classe Java é modificada e salva por um desenvolvedor, através do portal de desenvolvimento do CEJ, o Resin identifica tal modificação e automaticamente recompila a classe e carrega a nova versão na aplicação Web. Desta forma, o código do ambiente CEJ pode ser implementado independente da localização do desenvolvedor, bastando para isso um navegador *Web* e uma conexão com a Internet! A interface de navegação sobre os artefatos de desenvolvimento é ilustrada na figura a seguir:

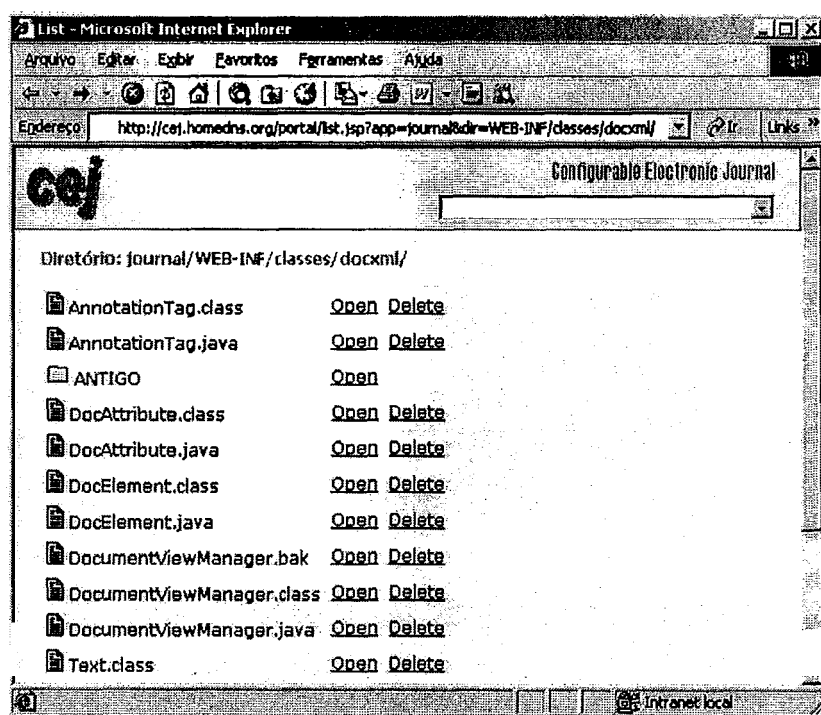


Figura 4.16 – Navegação sobre artefatos de desenvolvimento no ambiente CEJ

4.5.4.2 Acesso ao SGBD

A outra interface de apoio ao desenvolvimento à distância do ambiente CEJ oferece acesso ao SGBD acoplado ao ambiente CEJ. A interação com o SGBD é feita através da linguagem SQL ou de comandos específicos que podem ser trocados com o banco. Como requisitos para acessar uma instância do banco, o desenvolvedor deverá digitar o nome da instância e o nome de um usuário e senha para acessá-la.

É importante ressaltar que esta interface, ilustrada na figura 4.17, funciona com qualquer sistema gerenciador de bancos de dados que ofereça suporte a JDBC. No ambiente experimental, como já ressaltamos, foi utilizado o banco de dados MySQL.

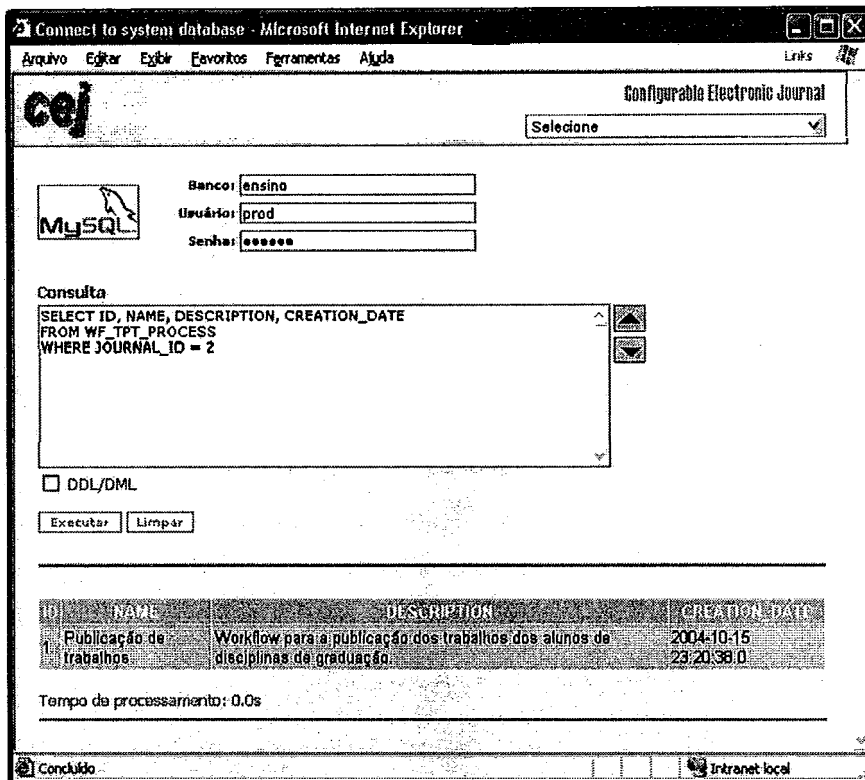


Figura 4.17 – Acesso ao SGBD pelo portal de desenvolvimento no ambiente CEJ

4.5.5 – Performance do ambiente CEJ

Por ser uma aplicação de lado servidor que roda Web, o ambiente CEJ deve considerar aspectos de performance e escalabilidade. Felizmente, o framework Struts, sob o qual o ambiente CEJ foi implementado, utiliza a API de Servlets, a qual roda em diversas linhas de execução (*threads*). O suporte a *threads* no ambiente Java é nativo, oferecendo excelente performance comparado a outros ambientes de desenvolvimento. Por outro lado, as classes do tipo *Action* do Struts são instanciadas somente uma vez, evitando o desperdício de recursos, sobretudo memória virtual.

Apesar de a infra-estrutura sob a qual o ambiente CEJ roda, em teoria, favorecer o aspecto de performance, alguns problemas de desempenho foram encontrados nos primeiros experimentos que foram feitos com o sistema. Particularmente, estes

problemas foram relacionados ao gerenciamento da memória virtual dedicada ao JVM (*Java Virtual Machine*), que é o processo que executa aplicações em Java em um sistema operacional.

Para solucionar tais problemas, algumas medidas foram necessárias para permitir que experimentos reais fossem viabilizados neste ambiente: 1) aumento da memória física do servidor que hospeda o ambiente CEJ; 2) diminuição do tempo máximo de inatividade de uma sessão de usuário – com isso os recursos ocupados por uma sessão passaram a ser liberados mais cedo, permitindo sua reutilização; 3) implementação do caso de uso “Encerrar seção”, dando ao usuário a possibilidade de antecipar a liberação dos recursos alocados à sua sessão; e 4) Otimização do código de alguns casos de uso e de algumas classes do sistema.

Dado que o gerenciamento da memória pela máquina virtual do Java tornou-se uma aspecto crítico, foi criada uma interface no ambiente CEJ para o acompanhamento da utilização da memória pela aplicação. Uma foto desta interface é exibida na figura 4.18:

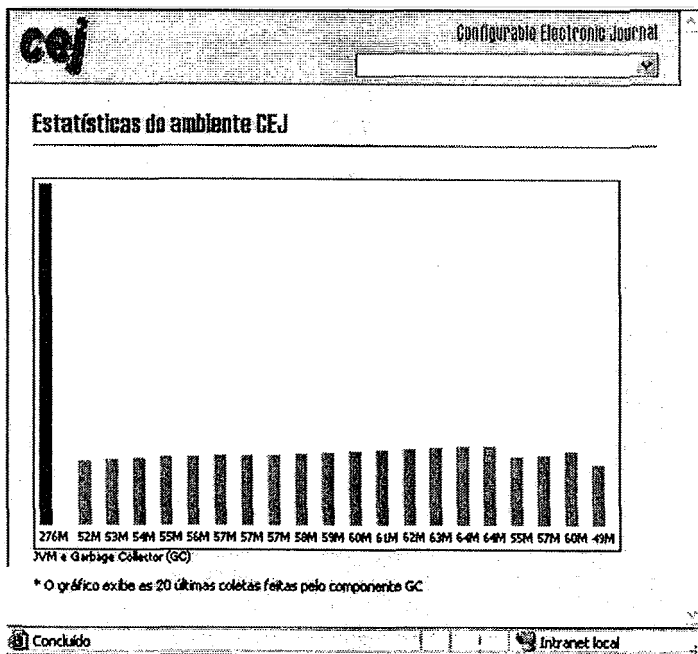


Figura 4.18 – Tela de “Estatísticas do ambiente CEJ”

Desempenho é também uma questão importante quando se trata de um *workflow* baseado na Web, mas os experimentos realizados com o protótipo do ambiente CEJ demonstraram que a solução encaminhada para o componente de *workflows* é plausível.

Capítulo 5

Estudos de caso com o ambiente CEJ

5.1 Considerações iniciais

Neste capítulo apresentaremos dois modelos de publicação científica que estivemos interessados em experimentar no ambiente CEJ. Estes modelos são também descritos em SCHNEIDER et al. (2004).

- 1) Uma revista de fluxo contínuo com publicação instantânea**
- 2) Uma revista baseada no convite com uma fase de proposta**

Para validar os conceitos propostos nesta monografia e implementados nesta versão experimental do ambiente CEJ, foram planejados dois experimentos “reais” com os modelos acima:

- I. O modelo (1) foi empregado na revisão colaborativa dos capítulos desta monografia.
- II. O modelo (2) foi utilizado em um experimento com duas turmas de graduação da UFRJ, no qual os alunos foram convidados a publicar trabalhos científicos sobre temas previamente escolhidos.

Na seção 5.2 descreveremos os modelos (1) e (2). Prosseguindo, as seções 5.3 e 5.4 descreverão, respectivamente, os experimentos I e II realizados com o ambiente CEJ. Finalmente, a seção 5.5 apresenta as principais conclusões obtidas sobre os experimentos.

5.2 Modelos propostos para a validação do ambiente CEJ

5.2.1 Modelo 1: Uma revista de fluxo contínuo com publicação *instantânea*

Neste modelo, os autores desenvolvem uma pesquisa, criam e editam os artigos, e então armazenam os mesmos em seus servidores pessoais na Web. Uma vez que o

artigo fica disponível na Web, leitores e pesquisadores podem acessá-lo livremente e lê-lo. Note, na figura 5.1, que a submissão do artigo pode ser feita tanto pelo autor quanto pelo leitor. O papel de “submissão” poderia ainda ser realizado pelo próprio editor da revista. Se o artigo é enviado com sucesso, ele é incorporado na revista através de um apontador (URL). Contanto que nenhuma revista detenha controle físico sobre o artigo, o artigo poderá ser publicado em diversas revistas ao mesmo tempo, conforme enfatizado no capítulo 2 da monografia.

Observe que, neste modelo, a publicação do artigo acontece imediatamente após a submissão - razão pela qual utilizamos o termo “publicação instantânea” - e acontece antes da fase de revisão pelos pares. O ciclo de publicação termina quando o editor formata o artigo, caso necessário, e o mesmo é retornado para atualização pelos autores. O modelo de fluxo contínuo é ilustrado na figura 5.1:

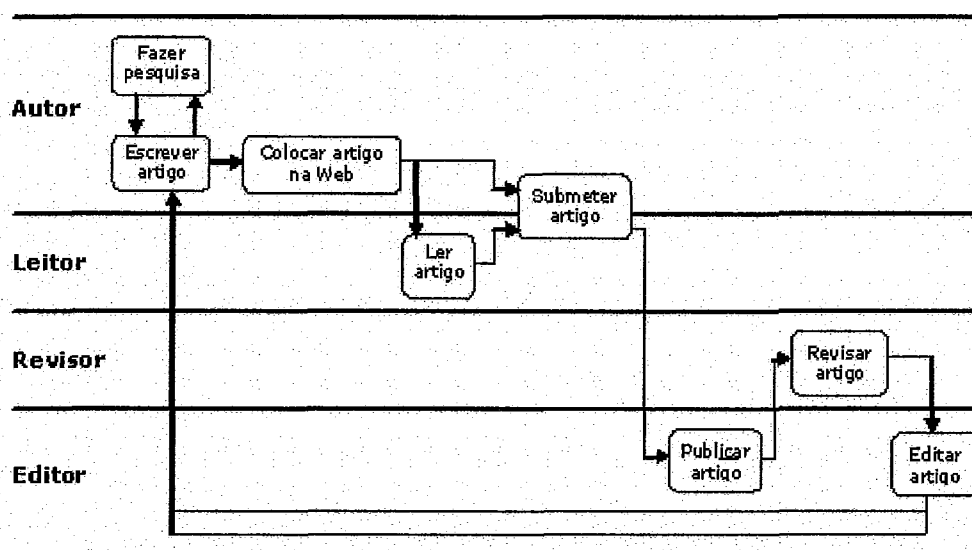


Figura 5.1 – O modelo de fluxo contínuo com publicação instantânea

Uma maneira de implementar este modelo no ambiente CEJ é definindo quatro fases: submissão, revisão, aprovação e atualização. Todas as quatro fases podem ser definidas simplesmente como fases elementares, que já existem de forma nativa no ambiente CEJ. A fase de submissão seria configurada para aceitar URLs, não sendo permitido o armazenamento físico dos artigos no servidor do CEJ. Na fase de revisão, a revista seria configurada para suportar revisão pelos pares *aberta*, e o sistema de

anotações configurado para permitir diversos tipos de contribuições e anotações. A fase de aprovação seria configurada para permitir que editores e revisores selecionassem as contribuições e pedidos de alteração por parte dos revisores e incorporá-los em um documento virtual, que ficaria disponível para o autor. Finalmente, na fase de atualização, um *deadline* seria configurado para edição pelos autores.

Existem muitas oportunidades para colaboração no modelo de publicação instantânea descrito aqui. Colaboração na forma de anotações, linhas de discussão e comentários pelos pares poderiam começar tão logo um artigo fosse submetido e publicado instantaneamente, e poderia ser realizada em paralelo com a atividade de revisão pelos pares. Contribuições poderiam continuar mesmo após as fases de revisão e atualização. Eventualmente, o artigo poderia ser estendido e um novo ciclo de revisão poderia ser realizado. Este, na verdade, é um modelo genérico, e muitas variações poderiam ser experimentadas incrementalmente.

Algumas considerações sobre o modelo (1) merecem ser mencionadas aqui. Primeiro, iniciativas de *self-archiving*, conforme comentado no capítulo 2, vêm ganhando força há mais de uma década, e desde 1994 têm sido defendidas pelo pesquisador Stevan Harnad, um dos mais proeminentes defensores dos repositórios de acesso aberto à literatura científica. O repositório da Física de Alta Energia, atribuído originalmente a Paul Ginsparg e discutido também no capítulo 2, é exemplo de uma implementação com sucesso dos modelos baseados no *self-archiving*.

Segundo, em contraste com o modelo tradicional, baseado no papel, a revisão pelos pares é realizada após a publicação no modelo de publicação instantânea, o que aumenta significativamente o potencial de redução no tempo do ciclo de publicação. Vimos que no ambiente eletrônico publicar é um ato reversível: é possível publicar um artigo e revisá-lo depois. O contra-argumento de que isso poderia levar a uma inundação de artigos de baixa qualidade assume incorretamente que o leitor não é capaz de distinguir entre artigos não-revisados e revisados. Além disso, tecnologias de filtragem podem ser utilizadas para eliminar sucessivamente os artigos indesejados. No ambiente CEJ, estampas são associadas aos documentos de forma a distinguir claramente os artigos revisados dos não-revisados. A interface também permite aos usuários acessar somente os artigos revisados, caso seja desejado.

5.2.2 Modelo 2: Uma revista baseada no convite com fase de proposta

Neste segundo modelo, o editor da revista identifica um artigo em potencial, baseando-se na leitura de um determinado artigo, e convida o autor do artigo para publicar outro. O autor escreve uma proposta, que é então revisada e negociada. Uma negociação bem-sucedida resulta na escrita e publicação do novo artigo.

Uma das principais vantagens desta abordagem é que praticamente todo artigo escrito vem a ser publicado. A desvantagem é que a seleção dos autores pode ser arbitrária e um autor desconhecido poderá nunca vir a ser candidato a publicar na revista. O modelo de publicação baseado no convite aqui descrito é ilustrado na figura 5.2:

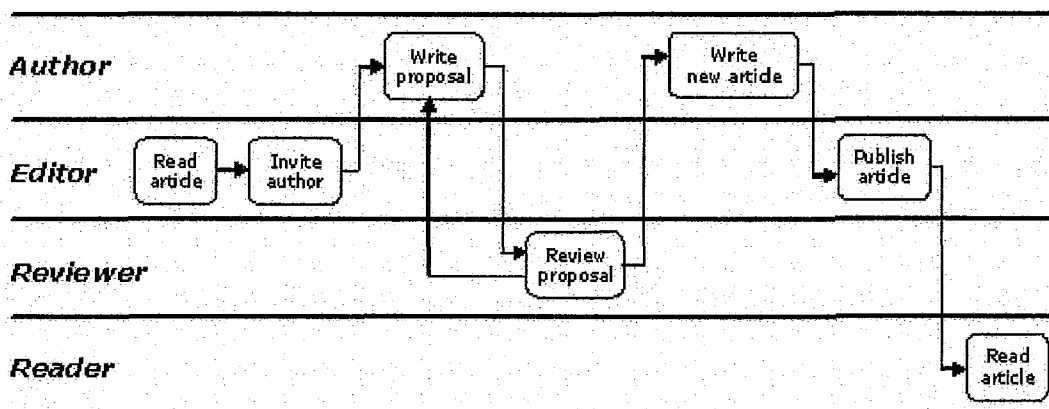


Figura 5.2 – Uma revista baseada no convite e com uma fase de proposta

A fim de implementar esse modelo no ambiente CEJ, podemos definir uma fase de *proposta*, composta por fases básicas de submissão, revisão e atualização. Uma segunda fase de submissão é também necessária para o novo artigo. Colaboração entre autores e revisores ocorrerá na fase de negociação, fazendo uso do sistema de anotações, e poderá continuar sendo feita sobre o novo artigo, após sua publicação.

É importante mencionar que as revistas tradicionais focam principalmente nos estágios finais do processo de pesquisa, por causa das limitações inerentes ao papel em suportar discussões férteis (HARS, 2003). A fase de proposta preenche uma grande lacuna das revistas tradicionais ao permitir que autores e revisores cooperem diretamente

nos estágios iniciais do *workflow* de publicação. Esta cooperação é proveitosa e benéfica tanto para os autores quanto para os publicadores. Ela facilita os autores na tarefa de produzir artigos que alcancem a audiência da revista. O outro lado da moeda é que tal cooperação facilita a identificação de um bom problema para pesquisa baseado em um trabalho anterior. Mesmo que a negociação de uma proposta não seja bem sucedida, a discussão é ainda válida na medida em que fornece um *feedback* construtivo aos autores da proposta, permitindo que eles caminhem em direção ao aprimoramento da pesquisa.

5.3 Descrição do experimento I

O primeiro experimento realizado com o ambiente CEJ teve como objetivo a revisão dos capítulos que compõem esta tese. Neste estudo de caso, que envolveu alunos e professores da UFRJ, foram sugeridas correções, melhorias e aprimoramentos no texto da monografia. Esta iniciativa pode ser vista como um primeiro passo rumo à implantação, na linha de Banco de Dados da COPPE, de um ambiente de suporte à geração, discussão e publicação de artigos e monografias, baseado na tecnologia CEJ.

Para suportar o experimento, foi criada no ambiente CEJ uma revista denominada **COPPEBD**. Seguindo a filosofia do modelo (1) descrito anteriormente, o autor desta monografia foi disponibilizando cada um de seus capítulos no servidor do ambiente CEJ empregado neste experimento. À medida que os capítulos eram disponibilizados, os revisores eram reportados através de mensagens automáticas da existência de um novo capítulo para revisão, dando início à revisão colaborativa.

O envolvimento de um número relativamente pequeno de revisores no experimento, aliado ao fato de que os participantes eram pesquisadores conhecidos do autor da monografia, possibilitou uma orientação personalizada aos membros envolvidos a respeito dos conceitos do ambiente CEJ - o que facilitou o entendimento de suas principais funcionalidades. Estas condições não existiram no segundo experimento, e isto será discutido na seção a seguir.

5.4 Descrição do experimento II

O segundo experimento realizado com o ambiente CEJ envolveu duas turmas de graduação da UFRJ – uma turma de *Engenharia de Software* do curso de Ciência da

Computação, e uma turma de *Inteligência Artificial* do curso de Engenharia Eletrônica. Os alunos das duas turmas foram divididos em *grupos de trabalho* e convidados a publicar trabalhos científicos sobre temas previamente escolhidos, sob autorização do professor. O modelo de publicação empregado, na verdade, representa uma pequena variação do modelo (2) descrito na seção anterior, envolvendo uma fase de negociação de uma proposta, seguida da confecção “definitiva” do documento. Nesta seção faremos uma breve descrição do experimento realizado com a turma de Engenharia de Software – composta por 40 estudantes - para a qual foi criada uma revista denominada **ES2004** no ambiente experimental do CEJ. Todos os alunos foram cadastrados na revista.

Tanto na fase de proposta quanto na fase de submissão dos trabalhos propriamente ditos, cada grupo revisou o trabalho de outro grupo. Estas revisões foram importantes para o processo de validação do ambiente CEJ pois possibilitaram a correção de inúmeros *bugs* encontrados no sistema, especialmente os *bugs* identificados no componente de navegação e anotações colaborativas.

O primeiro requisito para participar de um processo de publicação, conforme comentado, consistiu na escolha de um grupo de trabalho, o que é ilustrado na figura a seguir:

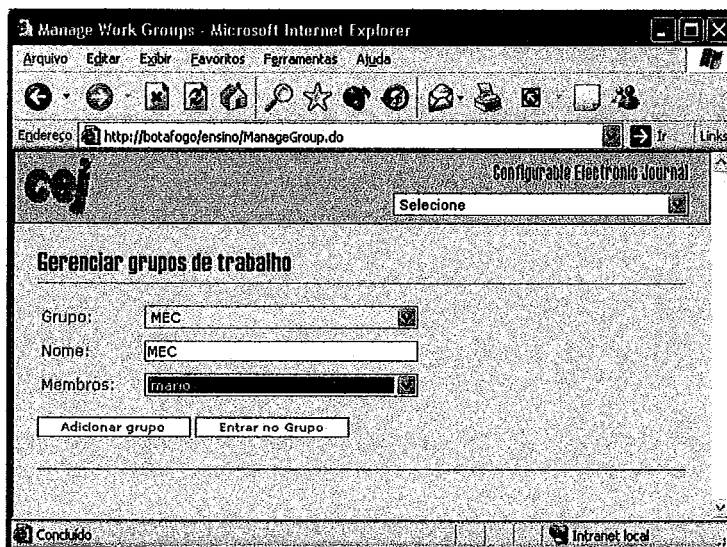


Figura 5.3 – Escolha de um grupo de trabalho

O outro requisito foi a escolha de um tema de trabalho. Estes temas foram pré-cadastrados pelo professor, mas aos alunos também foi permitido cadastrar novos, sob autorização do professor.

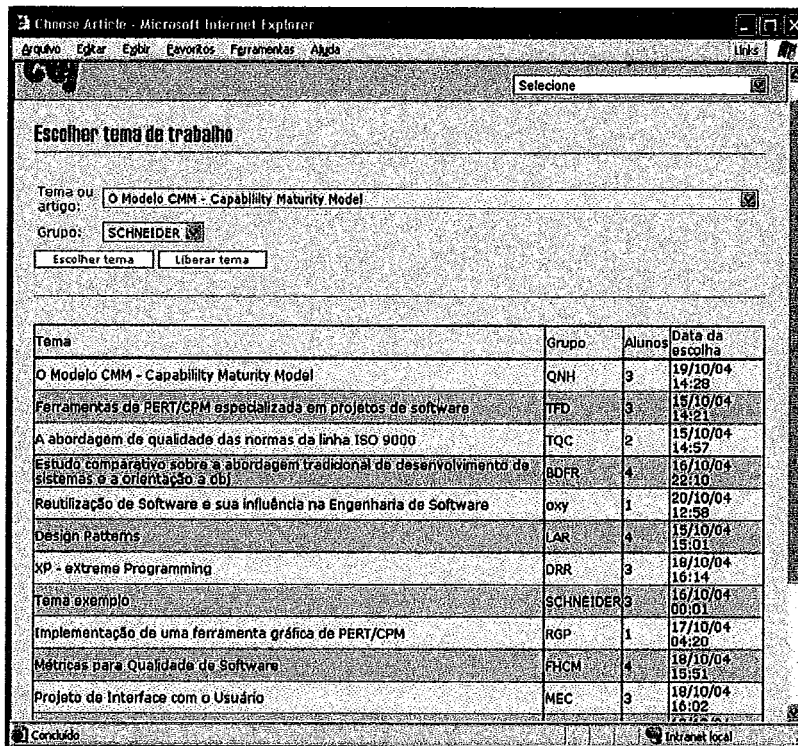


Figura 5.4 – Escolha do tema de trabalho

Com a reserva do tema de trabalho, os grupos deram início ao processo de publicação, cada qual criando uma instância do *template* de processo criado especialmente para este experimento. Já durante a fase de proposta muitos alunos reportaram dificuldades em utilizar algumas funcionalidades do sistema, o que reforçou a necessidade de que fossem disponibilizados manuais explicando como usar as tais funcionalidades, o que foi feito em seguida. Apesar destas dificuldades, a maioria dos grupos conseguiu trabalhar de forma plausível no ambiente, e mais de 20 documentos foram submetidos e revisados nesta revista, o que é ilustrado na figura 5.5:

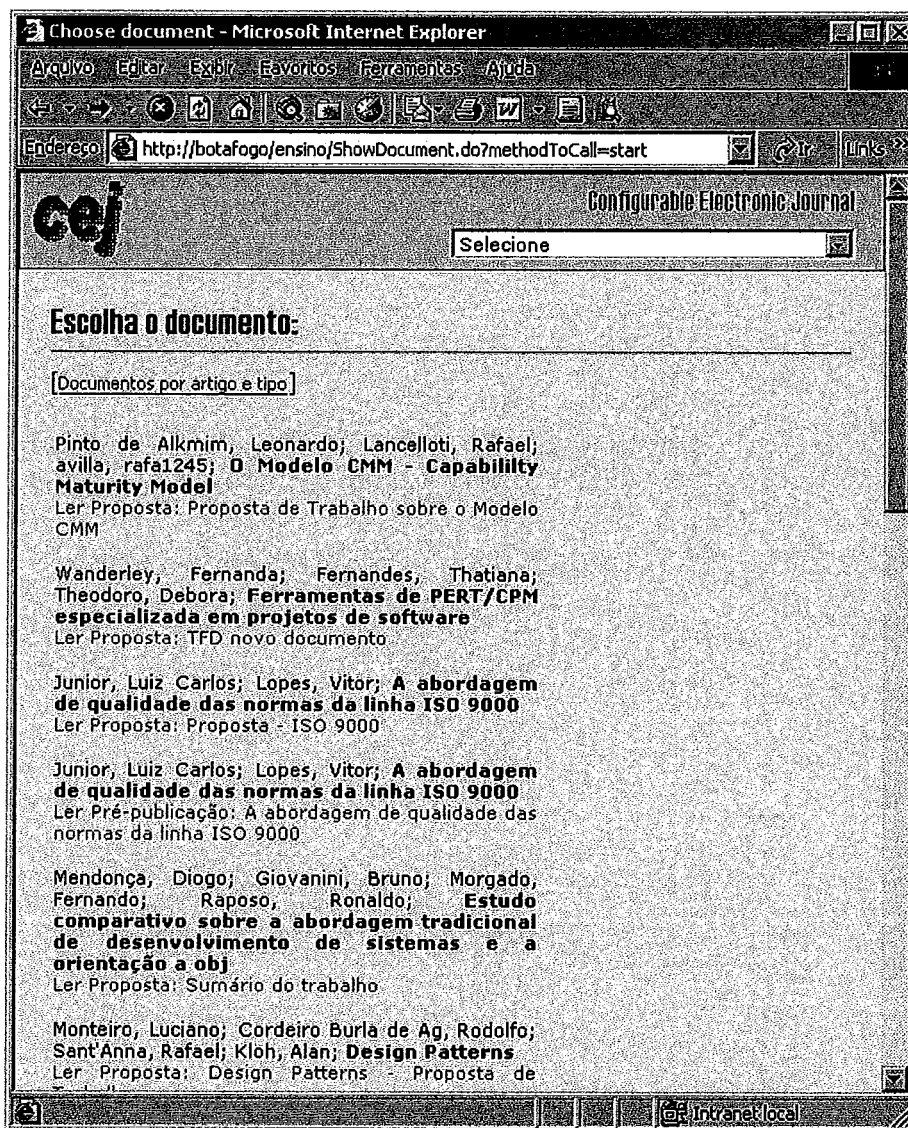


Figura 5.5 – Documentos submetidos à revista ES2004

Grande parte das dificuldades na utilização do sistema envolveu a submissão dos documentos e a atividade de revisão colaborativa. A submissão padrão dos documentos no ambiente CEJ, conforme proposto no capítulo 4, deve acontecer no formato XML DocBook, e muitos alunos não estavam familiarizados com tal formato. É bem verdade que o ambiente disponibiliza uma série de facilidades para apoiar a geração correta dos documentos neste formato – como o “Gerador de XML”, uma *parse* para verificar a correteude dos documentos, e uma funcionalidade de “*Preview*” que permite visualizar a exibição documento utilizando um dos *templates* padrão do sistema. Mesmo assim, o

esforço necessário para a submissão dos documentos foi ainda grande. Por outro lado, durante as fases de revisão dos documentos, muitos alunos não conseguiram explorar as funcionalidades de anotação e comentários pelos pares do ambiente, o que não invalida de maneira alguma o trabalho dos outros alunos que obtiveram êxito em suas atividades.

Todavia, um aspecto que deve ser enfatizado é que a grande maioria dos alunos mostrou desenvoltura na utilização da funcionalidade de participação no *workflow* de publicação. Acreditamos que isto se deve em parte à construção de uma interface simples e intuitiva.

5.5 Sumário e conclusões

Neste capítulo foram descritos dois modelos de publicação que foram alvo de simulação e experimentação no contexto deste trabalho de pesquisa. Conforme comentado, dois experimentos reais foram planejados com o objetivo de validar os conceitos propostos nesta monografia e a implementação do ambiente CEJ: duas revistas foram configuradas para suportar a publicação de trabalhos científicos de turmas da graduação da UFRJ, e uma terceira revista foi definida para possibilitar a revisão colaborativa desta monografia. Estes experimentos foram descritos nas seções 5.3 e 5.4.

Nossa avaliação dos experimentos realizados com o ambiente aqui proposto foi bastante positiva. Muitos dos conceitos centrais propostos aqui se mostraram plausíveis, como a submissão dos documentos no formato XML, a separação do conteúdo da apresentação, a tecnologia de *workflows* aplicada ao controle do processo de publicação, a revisão e o comentário pelos pares aberta e colaborativa, o suporte às anotações localizadas e às linhas de discussão centradas no documento, dentre outros.

Este trabalho de validação e experimentação dos conceitos apresentados nesta monografia será certamente estendido nos próximos meses, o que será discutido no capítulo seguinte.

Capítulo 6

Conclusões e trabalhos futuros

É chegada a hora de finalizarmos esta monografia e sintetizarmos as principais contribuições deste trabalho de pesquisa. Conforme comentado na introdução da monografia, um dos principais objetivos deste trabalho foi aplicar padrões e tecnologias abertas, sobretudo as da Web Semântica, incluindo-se aí a *eXtended Markup Language* (XML), e a tecnologia de anotações localizadas, na construção de um ambiente inovador para a definição e execução de processos de publicação flexíveis. Outros objetivos não menos importantes incluíram a experimentação de formas colaborativas de comunicação científica, e a investigação de novos papéis e atores no contexto de um processo de publicação científica. O principal fruto deste trabalho foi o nascimento do ambiente CEJ, que possibilita a construção de revistas eletrônicas flexíveis, onde os editores podem definir de forma customizável as atividades que vão compor um processo de publicação na revista, e onde os membros da revista podem participar de um processo de uma forma interativa e colaborativa.

Este trabalho iniciou-se com uma revisão extensa da literatura sobre as infra-estruturas de conhecimento científico no capítulo 2, incluindo as revistas científicas tradicionais, a principal entidade do modelo tradicional de publicação acadêmica. Prosseguindo, descrevemos as principais mudanças no processo de publicação que estão em curso em decorrência do advento da mídia eletrônica e da Internet. Esta descrição foi sucedida por uma revisão bibliográfica de alguns modelos emergentes propostos nos últimos anos por revistas eletrônicas que emergiram nos últimos anos. Uma caracterização das infra-estruturas de conhecimento científico, seguindo o *framework* de características proposto por HARS (2003), também foi realizada com o objetivo de levantar as principais bases do projeto do ambiente CEJ.

O capítulo 3 teve como principal objetivo uma análise comparativa das principais soluções de suporte a anotações colaborativas na Web, com o intuito de selecionarmos as melhores oportunidades para a construção do componente de navegação e anotações do CEJ. Finalmente, os capítulos 4 e 5 apresentaram, respectivamente, a descrição do

ambiente CEJ e a análise dos primeiros experimentos realizados com o protótipo deste ambiente.

Neste sentido concluímos que se encontra neste momento em curso um processo profundo de transformação nas formas de publicação e comunicação científicas. Algumas infra-estruturas têm caminhado no sentido da evolução da revisão pelos pares de sua forma originalmente fechada para a forma aberta, em que os autores do artigo conhecem os revisores e vice-versa. Conforme discutimos, inúmeras críticas ao processo tradicional motivaram a confecção de novos modelos, e muitas revistas estão experimentando atualmente essas novas formas, e muitos casos de sucesso têm sido relatados na literatura sobre o assunto. Dentre os casos de sucesso, analisamos o exemplo da Física de Alta Energia, as iniciativas de Stevan Harnad na área de Psicologia e áreas afins, e algumas revistas na área da Medicina, entre as quais a BMJ se destaca como um exemplo notório. Como o momento é de transição, não nos arriscamos em afirmar que nenhum destes modelos seja representativo da nova realidade que apenas começa a ser desenhada, mas uma análise das características relevantes e inevitáveis pôde ser feita com risco mínimo.

Apesar das iniciativas apresentadas, muitos problemas ainda permanecem abertos, sem perspectivas de solução. De modo geral, mesmo as revistas emergentes ainda se prendem a resquícios inerentes ao modelo de Gutenberg; em alguns casos, apesar de a revisão pelos pares ter evoluído para a modalidade aberta, o processo continua produzindo uma decisão única de publicar ou não um determinado artigo.

O intervalo entre a submissão de um artigo e sua publicação ainda é grande na maior parte dos casos (HARS, 2003), dificultando o acesso rápido à informação, tão importante para o progresso das ciências. Por outro lado, comentários pelos pares pós-publicação, adotado por algumas revistas nos últimos anos, parece ser um valioso suplemento à revisão tradicional, mas definitivamente não é um substituto para ela.

Diversos problemas foram analisados, equacionados e melhorados ou resolvidos neste trabalho de pesquisa, e o capítulo 4 apresentou as principais soluções encaminhadas. Cabe agora fazer uma análise das principais contribuições desta pesquisa, o que será feito a seguir.

6.1 Contribuições desta dissertação

Uma das possibilidades praticamente não explorada pelas atuais infra-estruturas de conhecimento científico – de acordo com HARS (2003) - abordada neste trabalho, foi a do emprego de uma máquina de *workflow* no ambiente de publicação, que possibilita a automatização das diversas atividades que compõem um processo de publicação, incluindo a submissão de documentos, a revisão e comentários pelos pares, a delegação de atividades etc. Dois aspectos importantes da máquina de *workflow* do CEJ são o suporte a o *gerenciamento dinâmico dos processos de publicação*, que permite que se modifique um processo de publicação, mesmo que algumas instâncias deste processo estejam em execução, e *as interações pela Web*, como acontece com o ambiente como um todo, que possibilita a participação dos membros de uma revista em um processo de publicação independente da localização do usuário e do sistema operacional utilizado, sendo suficientes uma conexão com a Internet e um navegador *Web*. No tocante às atividades de revisão e comentários pelos pares, o ambiente construído suporta a configuração de revisões abertas ou fechadas, e diversos tipos de interação podem ser empregados nestas atividades, com o apoio dos componentes de *navegação e anotação*, e do componente de *workflow customizável*, possibilitando ainda a definição de diversas fases, em sintonia com a proposta de revisão continuada proposta por ODLYZKO (2003).

Uma contribuição importante deste trabalho foi o emprego de tecnologias abertas, sobretudo as tecnologias propostas pela W3C para a Web, na construção de um ambiente para a definição (e execução) de processos de publicação científica. Isto possibilita a extensão futura deste trabalho não apenas pelo próprio autor como por outros grupos de pesquisa. Aqui vale ressaltar a importância da utilização do *paradigma orientado a objetos*, da adoção do *framework* Struts no projeto e implementação, e da construção do *portal de desenvolvimento do CEJ*, que garantiram um desenvolvimento flexível deste ambiente, o que possibilitou uma reutilização em escalas maiores e favoreceu sobremaneira os trabalhos futuros.

Uma proposta inovadora embutida no ambiente CEJ foi a de separar o conteúdo de um documento de sua apresentação, através da tecnologia XML e do uso de *templates* de exibição, que podem ser compartilhados também no formato XML.

Uma outra abordagem explorada nesta pesquisa foi a de anotações colaborativas sobre documentos XML, utilizando a filosofia dos sistemas baseados em uma aplicação *proxy*, cuja principal responsabilidade é realizar a junção do documento com suas anotações.

Outras contribuições deste trabalho incluem o suporte ao armazenamento flexível dos documentos em servidores públicos, sem exigir que estejam localizados no servidor do CEJ; a geração dinâmica do conteúdo de um documento, customizada aos usuários e sensível ao contexto da atividade de publicação; a busca de documentos e anotações com maior contexto e precisão, já que os elementos de um documento ganham uma semântica especial; e a geração automática de uma versão para impressão dos documentos, no formato PDF, a partir do documento original no formato XML.

Ao longo deste trabalho de pesquisa o autor foi coletando diversas opiniões, boatos e mentiras encontrados na *Web*, e até argumentos inconsistentes ou contraditórios defendidos pelas próprias revistas científicas tradicionais, no tocante à discussão da publicação científica eletrônica. Estas informações foram sintetizadas pelo próprio autor em quatro mitos, que serão apresentados a seguir e também se encontram disponíveis no portal do ambiente CEJ.

6.1.1 Quatro mitos da publicação científica

(1) A revisão pelos pares tradicional, realizada pelas grandes revistas e conhecida como *peer-review*, é a única forma "segura" de validar um artigo científico na era da Internet e da mídia eletrônica.

Mito! O *peer-review* tradicional é um processo realizado de forma fechada e uma única vez, que produz uma decisão binária de publicar ou não um artigo. Existem outras formas também "seguras" e até mais eficientes de validar um artigo científico, como comentários pelos pares e revisões abertas, continuadas e multidimensionais, onde a divisão binária entre o que foi publicado e o que não foi publicado pode ser substituída por um espectro.

(2) A atividade de revisão de um artigo científico deve vir sempre antes da sua publicação.

Mito! No ambiente eletrônico, é possível publicar um artigo de forma "instantânea", isto é, imediatamente após sua submissão e enquadramento no escopo e objetivos da revista científica. O argumento contrário, de que a revisão precisa vir sempre antes da publicação, assume incorretamente que os leitores não são capazes de distinguir entre um artigo que ainda não foi revisado, e um artigo já revisado. A tecnologia de filtros eletrônicos deve ser utilizada para garantir que os leitores acessem somente os artigos que sejam de seu interesse.

(3) O PDF (*Adobe Portable Document Format*) é o padrão mais adequado ao armazenamento e navegação sobre artigos científicos na Web, razão pela qual ele é o formato adotado de forma predominante pelas revistas científicas.

Mito! O padrão PDF "mistura" o conteúdo de um artigo com a sua apresentação, impedindo que leitores com perfis distintos possam visualizar um documento de forma customizada. Este formato não se compatibiliza com as principais tecnologias da Web, como a manipulação de seu conteúdo através do modelo de objetos do documento (DOM). Mais adequado é o formato XML, que além de possibilitar tal separação integra-se perfeitamente com o ambiente Web.

(4) Com a migração das revistas científicas para a Web nos últimos anos, um novo modelo de publicação baseado na mídia eletrônica foi implementado com sucesso, resolvendo os principais problemas inerentes ao modelo tradicional.

Mito! Preocupações a respeito de possíveis prejuízos ao sistema tradicional de revisão pelos pares estão retardando o passo da mudança. Apesar do aparente progresso demonstrado por algumas revistas emergentes, o modelo de publicação científica baseado no papel, e centrado nas grandes revistas, permanece essencialmente "clonado" no ambiente eletrônico. Se levarmos em conta o universo de 24.000 revistas científicas revisadas publicadas no mundo inteiro (HARNAD, 2004), os casos inovadores constituem ainda a exceção, e não a regra.

6.2 Trabalhos futuros

O trabalho de pesquisa desenvolvido nesta monografia, que resultou na construção do ambiente CEJ, não se encerra no momento da defesa. É de desejo do autor e de seus colaboradores que este trabalho seja estendido e constantemente aprimorado.

Como trabalhos futuros, está prevista uma nova fase de desenvolvimento deste ambiente. Entre as melhorias vislumbradas podemos citar o suporte ao versionamento de documentos, problema que não foi equacionado neste trabalho de pesquisa; a extensão da API do componente de navegação e anotações, dando maior suporte à interatividade e à geração das âncoras das anotações localizadas; o suporte do sistema no lado *cliente* a outros navegadores *Web*, como aqueles baseados no navegador aberto e gratuito *Mozilla*, já que o protótipo atual focou no navegador oficial da Microsoft – o *Internet Explorer*; a funcionalidade do "Gerador de XML", que permite a conversão de um documento HTML para o formato XML, livre de estilos e formatações, precisaria também ser aprimorada. Estas extensões mais imediatas do presente trabalho terão como objetivo a implantação do ambiente CEJ nos grupos de pesquisa da linha de Banco de Dados da COPPE, suportando a revisão de teses, periódicos e apoiando a confecção de artigos dos alunos de mestrado e doutorado. É também um desejo do autor que este ambiente seja experimentado, já no ano de 2005, no apoio à realização de alguma conferência nacional ou ao processo de publicação de alguma revista científica real.

Outras extensões do presente trabalho incluem o suporte a outros tipos de processos de publicação não suportados pela atual infra-estrutura do componente de *workflow* do CEJ; a investigação da abordagem de Sistemas Orientados a Serviços (*Web Services*) como extensão da arquitetura do CEJ, permitindo a integração com outras infra-estruturas de conhecimento científicas disponíveis na *Web*. Sistemas Orientados a Serviços constituem uma nova abordagem arquitetônica baseada na noção de se construir aplicações através da descoberta e da interligação orquestrada (OLIVEIRA, 2003).

O portal de desenvolvimento do CEJ também poderia ser explorado de forma mais consistente em desenvolvimentos futuros. O argumento de que o desenvolvimento de um sistema pela *Web* é ineficiente pode hoje ser rebatido em função da popularização

das tecnologias de *banda larga*. Uma outra possibilidade é o desenvolvimento do sistema através da *Intranet*. A grande vantagem deste tipo de desenvolvimento é que só exige um navegador e a conexão com alguma rede interligada ao servidor de desenvolvimento do CEJ.

A tradução dos *templates* de processos e exibição de documentos para o formato XML ainda se encontra em fase de testes. Finalmente, o suporte à *internacionalização*, previsto na proposta do ambiente CEJ, não foi implementado no primeiro protótipo, o que deverá ser feito nos próximos meses.

6.3 Considerações finais

Para finalizar, alguns comentários dedicados aos entusiastas, àqueles que em sintonia com o autor desta monografia anseiam por um modelo de publicação científica realmente novo, livre de qualquer resquício da mídia de Gutenberg que tanto progresso trouxe à humanidade, mas que aos poucos vem sendo substituída pela mídia eletrônica. Um modelo que beneficie os pesquisadores, e não apresente como principal objetivo trazer lucros aos editores das grandes revistas.

Uma nova era para a publicação científica está se consolidando, com a convergência da mídia eletrônica, da Internet, e da computação baseada nos padrões abertos. Conforme destacamos no capítulo 2 desta monografia, nos últimos anos ganhou força um movimento em defesa dos repositórios abertos, conhecido como *Open Archives Initiative* (OAI), tendo como principal objetivo criar um repositório único virtual de literatura científica. A concretização desses repositórios distribuídos e interoperáveis beneficiaria as instituições de pesquisa de três maneiras. Primeiro, maximizaria a visibilidade e o impacto da pesquisa revisada da instituição. Segundo, maximizaria o acesso pelos pesquisadores ao resultado da pesquisa revisada de todas as outras instituições. Terceiro, as instituições de pesquisa poderiam deflagrar a transição do modelo antigo de publicação ao modelo OAI, agilizando a derrocada do primeiro.

No momento em que esta monografia era finalizada, mais uma ferramenta gratuita de apoio à pesquisa científica era disponibilizada na Web, com características semelhantes ao serviço *Citeseer* citado no capítulo 2: o motor de busca de artigos do Google: <http://scholar.google.com/>. A ferramenta permite a busca de artigos disponíveis

na Web nos principais formatos de publicação (PDF, PS e HTML) e implementa apontadores bidirecionais, isto é, dado um artigo, o sistema indica que artigos encontrados na Web citam este artigo, além da tradicional lista de referências oferecida tradicionalmente. Ainda em sua versão beta, esta máquina de busca certamente ainda será aprimorada e evoluída, e poderá representar mais um passo importante rumo ao acesso gratuito da literatura científica na *World Wide Web*, e à sedimentação da idéia de que o artigo que não estiver disponível de forma gratuita na Web, estará invisível (LAWRENCE, 2001).

Deste modo, ao chegarmos ao final desta monografia, somos forçados a reconhecer que apenas arranhamos a superfície de um enorme monolito. De um ponto de vista mais positivo, conseguimos definir algumas das características que deverão estar de alguma forma presentes nas futuras infra-estruturas de conhecimento científico. Neste sentido, acreditamos que o ambiente CEJ possa representar a essência das características que imaginávamos para tais infra-estruturas no início deste trabalho de pesquisa.

Se os conceitos e propostas apresentados aqui serão representativos dos futuros modelos de publicação, só o tempo dirá. Uma coisa, porém, parece certa: estes três gigantes que são a **mídia eletrônica**, a **Internet**, e a **computação baseada em padrões abertos**, cada um com sua contribuição, marcarão a história da publicação científica no século XXI.

Referências Bibliográficas

- ALLEN, R., 2001, "Workflow: An Introduction". Disponível em: http://www.wfmc.org/standards/docs/Workflow_An_Introduction.pdf. Acesso em: 5 ago. 2003.
- ARAÚJO, R. M., BORGES, M. R. S., 2001, "Sistemas de Workflow". *XX Jornada de Atualização em Informática*, Congresso da SBC, Rio de Janeiro.
- ARMSTRONG, T.S., 1997, "Peer review for journals: Evidence on quality control, fairness and innovation", *Science and Engineering Ethics*, vol. 3, pp. 63-84.
- BAILEY, C. W., 2001, "Scholarly Electronic Publishing Bibliography". Disponível em: <http://info.lib.uh.edu/sepb/sepb.html>. Acesso em: 28 jan. 2003.
- BARUA, A., CHELLAPPA, R., WHINSTON, A.B., 1995, "Creating a collaboratory in cyberspace: Theoretical foundations and implementation", *Journal of Organizational Computing*, v. 5, n. 4, pp.417-442.
- BERNERS-LEE, T., et al., 2001, "The Semantic Web". Disponível em: <http://www.sciam.com/article.cfm?articleID=00048144-10D2-1C70-84A9809EC588EF21>. Acesso em: 23 ago. 2003.
- BOYCE, Peter B., 2000, "For Better or Worse: Preprint Servers Are Here to Stay", *College and Research Libraries News*, vol. 61, no. 5, pp. 404-407.
- BRIN, S., PAGE, L., 1998, "The Anatomy of a Large-Scale Hypertextual Web Search Engine". Disponível em: <http://www.db.stanford.edu/~backrub/google.html> Acesso em: 19 dez. 2004.
- BRUSH, A. J. B., 2002, "Annotating Digital Documents for Asynchronous Collaboration", *Technical Report*, Dept. of Computer Science and Engineering.
- BUSH, V., 1945, "As We May Think", *Atlantic Monthly*, vol. 176, no. 1, pp. 101-108.
- CASTRO, M. , 2003, *DECISIO: Um sistema de apoio à decisão para o planejamento ambiental*, Tese de M. Sc., COPPE / UFRJ, Rio de Janeiro.
- CECI, S. J. & PETERS, D. P., 1984, "How Blind is Blind Review?", *American Psychologist*, n. 39, pp. 1491-1494.
- CHAFFEY, D., 1998, *Groupware, Workflow and Intranets: Reengineering the enterprise with collaborative software*, Butterworth-Heinemann.

- CYCORP, 1999, "Background on the Cyc Technology". Disponível em: <http://www.cyc.com/tech.html>. Acesso em 18 jun. 1999.
- DAVIS, J., 1996, "CoNote - small group annotation experiment". Disponível em: <http://dri.cornell.edu/pub/davis/Annotation/annotation.html>. Acesso em: 20 jul. 1998.
- DENOUE, L., 1999, "Adding Metadata to improve retrieval: Yet Another Web Annotation System". Disponível em: <http://www.univ.savoie.fr/labos/syscom/Laurent.Denoue/yawas.doc>. Acesso em: 19 dez. 2004.
- DENOUE, L., 2002, "Annotations in the Wild". Disponível em: http://www.fxpal.com/people/denoue/publications/annotations_in_the_wild.pdf. Acesso em: 19 dez. 2004.
- ELLIS, C. A., GIBBS, S. J., REIN, G. L., 1991, "Groupware: Some issues and experiences". *Communications of the ACM*, v. 34, n. 1, pp. 38-58.
- GINSPARG, P., 1996, "Winners and losers in the global research village", *UNESCO conference*, Fevereiro, 1996.
- GLOGOFF, S., 1988, "Reviewing the gatekeepers: A survey of referees of library journals", *Journal of the American Society for Information Science*, v. 39, pp.400-407.
- GRAMLICH, C. W., 1997, "Public Annotation Systems". Disponível em: <http://playground.sun.com/~gramlich/1994/annotate/index.html>. Acesso em: 12 ago. 2004.
- GUTWIN, C. & GREENBERG, S., 1998, "Effects of Awareness Support on Groupware Usability". *Proceedings of the Conference on Human Factors in Computing Systems (CHI'98)*, pp.511-518, Los Angeles, California, USA.
- GUZDIAL, M., 1997a, "Anchored collaborative learning environments". Disponível em: <http://www.cc.gatech.edu/gvu/edtech/CaMILE.html>. Acesso em: 12 ago. 2004.
- GUZDIAL, M., 1997b, "Information ecology of collaborations in educational settings: Influence of tool". *Proceedings of the 2 nd International Conference on Computer Supported Collaborative Learning (CSCL '97)*, pp. 83-90, Toronto.
- HAMILTON, D.P., 1990, "Publishing by – and for? – the numbers?", *Science* 250, pp.1331-1332.
- HARNAD, S., 1990, "Scholarly Skywriting and the Prepublication Continuum of Scientific Inquiry", *Psychological Science*, vol. 1, pp. 342-343.

- HARNAD, S., 1991, "Post-Gutenberg Galaxy: The Fourth Revolution in the Means of Production of Knowledge". *R. D. Mason (ed.) Computer Conferencing: The Last Word*.
- HARNAD, S., 1996, "Implementing Peer Review on the Net: Scientific Quality Control in Scholarly Electronic Journals". *Scholarly Publishing: The electronic frontier* (Peek, R. P. and Newby, G. B., eds) Cambridge, MA: The MIT Press, pp.103-118.
- HARNAD, S., 1998, "The invisible hand of peer review", *Nature*. Disponível em: <http://helix.nature.com/webmatters/invisible/invisible.html>. Acesso em: 5 nov. 1998.
- HARNAD, S., 2001a, "The self-archiving initiative", *Nature*, n. 410, pp. 1024-1025, 2001.
- HARNAD, S., 2001b, "AAAS's Response: Too Little, Too Late", *Science dEbate*. Disponível em: <http://cogprints.org/1641/00/science2.htm>. Acesso em: 19 dez. 2004.
- HARNAD, S., 2003, "Electronic Preprints and Postprints", *Encyclopedia of Library and Information Science*, Marcel Dekker, Inc.
- HARNAD, S., 2004, "Let All Knowledge Be Free That Wants to be Free". Disponível em: <http://www.ecs.soton.ac.uk/~harnad/Temp/gazette.html>. Acesso em: 5 nov. 2004.
- HARS, A., 2003, *From Publishing to Knowledge Networks*, Springer, Germany.
- HENDRICKSEN, C., 1999, "Asynchronous Collaboration by Document Review". Disponível em: <http://students.washington.edu/veritas/papers/DRpaper/basedoc.html>. Acesso em: 3 nov. 2004.
- HENDRICKSEN, C., 2002, *The Research Web: Asynchronous Collaboration in Social Scientific Research*, Tese Ph.D., University of Washington.
- HILGARTNER, S., 1995, "Biomolecular databases – New communications regimes for biology?", *Science communication*, vol. 17, no. 2, pp. 240-263.
- ITEXT, 2004, "iText: a Free Java-PDF library". Disponível em: <http://www.lowagie.com/iText/>. Acesso em: 19 dez. 2004.
- JOHNSON, R., 1998, "Designing Reusable classes". *Journal of Object-Oriented Programming*, vol. 1, no. 2, pp. 22–35.
- KAHAN, J. et al., 2002, "Annotea: an open RDF infrastructure for shared Web annotations", *Proceedings of the WWW10 International Conference*, Hong Kong, 2001.

- KENNEDY, É., 2003, "The Academic Peer Review Publication Process".
- KHOSHAFIAN, S., BUCKIEWICZ, M., 1995, *Introduction to Groupware, Workflow, and Workgroup Computing*. New York: John Wiley & Sons.
- KLING, R., FORTUNA, J., KING, A., 2001, *The real stakes of virtual publishing: The transformation of E-Biomed into PubMed Central*, Indiana University, Center for Social Informatics.
- KOBIELUS, J., 1997, *Workflow Strategies*. New York, IDG Books.
- KOULOPOULOS, M., 1995, *The Workflow Imperative. Building Real World Business Solutions*. Boston, John Wiley & Sons, Inc.
- LAFOLLETTE, M.C., 1992, *Stealing into print*, BERKELY.
- LAWRENCE, S., 2001, "Online or invisible?", *Nature*, vol. 411, no. 6837, pp. 521.
- LUCE, R., 2000, "The Open Archives Initiative: Interoperable, Interdisciplinary Author Self-archiving Comes of Age", *Serials Review and NASIG Conference Proceedings*, San Diego, USA.
- MANN, C. E., 2001, "Electronic paper turns the page", *Technology Review*, vol 104, no 2, pp. 42-48.
- MCKNIGHT, C., DILLON, A., SHACKEL, B., 1996, "The electronic journal and its implications for the electronic library", *Harrison*, pp. 351-368.
- MINTZER, F., BRAUDAWAY, G., BELL, A, 1998, "Opportunities for watermarking standards", *Communications of ACM*, vol. 41, no. 7, pp. 57-64.
- MORAIS, E. , 2003, *SDOC: Gerência de documentos e conteúdos científicos*, Tese de M. Sc., COPPE / UFRJ, Rio de Janeiro.
- O'CONNOR, M., 1986, *How to copyedit scientific books and journals*, Philadelphia, ISI Press.
- ODLYZKO, A. M., 1994, "Tragic loss or good riddance? The impending demise of traditional scholarly journals", *Journal of Human-Computer Studies*, vol 42.
- ODLYZKO, A. M., 1997, "The slow evolution of electronic publishing". *Electronic Publishing - New Models and Opportunities*, A. J. Meadows and F. Rowland, eds., ICC Press, pp. 4-18, 1997.
- ODLYZKO, A., 1999, "The evolution of electronic scholarly communication", *Thirty-second Annual Meeting of Assoc. Earth Science Editors*.

- ODLYZKO, A., 2001, "The rapid evolution of scholarly communication". *Proceedings of the 1999 PEAK conference*.
- ODLYZKO, A., 2003, "Peer and non-peer review". Disponível em: <http://www.dtc.umn.edu/~odlyzko/doc/peer.review.html>. Acesso em: 5 nov. 2004.
- OKERSON, A., 1996, "University libraries and scholarly communication", *Peek & Newby*, pp.181-199.
- OLIVEIRA, J., 2003, *EPISTHEME: Um Ambiente de Gestão de Conhecimento Científico*. Tese de M.Sc., COPPE/UFRJ, Rio de Janeiro, RJ, Brasil.
- POLTROCK, S. E. & GRUDIN, J., 1997, "CSCW, Groupware and Workflow: Experiences, State of Art and Future Trends". *Proceedings of the International Conference on Human-Computer Interaction (INTERACT'1997)*, pp. 661-662, Sydney, Australia.
- RÖSCHEISEN, M. & MOGENSEN, C. & WINOGRAD, T., 1995, "Shared Web Annotations as a Platform for Third-Party Value-Added Information Providers: Architecture, Protocols, and Usage Examples". Disponível em: <http://www-diglib.stanford.edu/diglib/pub/reports/commentor.html>. Acesso em: 23 ago. 2003.
- ROWLAND, F., 1996, "The need for management of electronic journals", *PEEK, R. P. e NEWBY, G. B. (eds.). Scholarly publishing: the electronic frontier*. Cambridge, MA: MIT Press, 1996, pp. 243 –250.
- ROWLAND, F., 2002, "The peer-review process", *Learned Publishing*, v. 15, n. 4, pp. 247-258.
- SCHNEIDER, D., SOUZA, J., XEXÉO, G., MEDEIROS, S., 2004, "Configurable Electronic Journal: Towards Flexible Scientific Knowledge Infrastructures". *Proceedings of the 8th International Conference on Computer Supported Cooperative Work in Design*, vol. 2, pp. 227- 232.
- SHUM, S. & SUMNER, T., 1996, "From Documents to Discourse: Shifting Conceptions of Scholarly Publishing". *Proceedings of CHI 98: Human Factors in Computing Systems*. ACM Press, New York.
- SHUM, S. & SUMNER, T., 2001, "JIME: An Interactive Journal for Interactive Media". *First Monday*, vol. 6, no. 2. Disponível em: http://www.firstmonday.dk/issues/issue6_2/buckingham_shum/index.html. Acesso em: 29 ago. 2003.
- SIMON, R.J., FYFE, J.J., 1994, *Editors as gatekeepers: Getting published in the social sciences*, BOSTON.
- SMITH, R., 1999a, "Opening up BMJ peer review", *BMJ*, v. 318, pp. 23-27.

- SMITH, J. W. T., 1999b, "The Deconstructed Journal – A new model for Academic Publishing", *Learning Publishing*, v. 12, n. 2.
- STORY, G.A., O'GORMAN, L., FOX, D., SCHAPER, L., 1992, "The Right-Pages image-based electronic library for alerting and browsing", *IEEE Computer*, v. 25, n. 9, pp.17-26.
- SUMNER, T. & SHUM, S. B., 1996, "Open Peer Review and Argumentation: Loosening the Paper Chains on Journals", *Ariadne*.
- TENOPIR, C., KING, D., 2000, *Towards electronic journals: Realities for scientists, librarians and publishers*, Washington, DC: Special Libraries Association.
- VALADARES, A., 2001, *Modelagem de processos para implementação de workflow: uma avaliação crítica*, Tese de M. Sc., COPPE / UFRJ, Rio de Janeiro.
- VARIAN, H., 1997, "The future of electronic journals", *Proceedings of the conference on Scholarly Communication and Technologies*.
- VASUDEVAN, V. & PALMER, M., 1999, "On Web Annotations: Promises and Pitfalls of Current Web Infrastructure". *Proceedings of the 32nd Hawaii International Conference on Systems Sciences*, Maui, Hawaii.
- WALKER, T., 1998, "Free Internet Access to Traditional Journals", *American Scientist*, vol. 86, no. 5, 1998. Disponível em: <http://www.sigmaxi.org/amsci/articles/98articles/walker.html>. Acesso em: 27 jun. 2001.
- WATSON, R., 1994, "Creating and sustaining a global community of scholars", *MISQ*, v. 18, n. 3, pp.228-231.
- WHITTINGTON, C. D., 1996, "Annotation systems: A review".
- WILLIAMSON, A., 2002, "What Happens to Peer Review?" *ALPSP International Learned Journals Seminar*, London, UK, 12 April 2002.
- WILSON, J.D., 1978, "Peer-review and publication: Presidential address for the 70st annual meeting of the American Society for Clinical Investigation", *Journal of Clinical Investigation*, v. 67, pp.1697-1701.
- W3C DOM WORKING GROUP, 1998, *Document Object Model (DOM) level 1 Specification, version 1.0, W3C recommendation*, Technical report, World Wide Web Consortium. Disponível em: <http://www.w3.org/TR/REC-DOM-Level-1>. Acesso em: 23 ago. 2003.
- YEE, K. P., 1999, "What's a mediator?" Disponível em: <http://zesty.ca/mediator.html>. Acesso em: 23 ago. 2003.

Apêndice A

Framework de características proposto por HARS (2003)

Tabela A.1 – Características das infra-estruturas do conhecimento científico (HARS, 2003)

Tipo		Alternativas							
Modo de interação	Tipos de interação suportados	Leitura e navegação		Anotação e comentários		Revisão	Autoria		
	Anotações online	Não	Sim, restrita a um grupo		Sim, por qualquer leitor		Sim, de forma anônima		
	Exibição das anotações	Não	Sim, no próprio documento		Sim, apontadas pelo documento		Sim, em outra mídia (ex: fórum)		
	Submissão online	Não	Via email		Via FTP	Via http	Criação e edição online		
	Momento de publicação	Na submissão		Na aceitação		Depois da aceitação ("Issue" da revista)			
	Atualização das contribuições	Não (contribuições imutáveis)		Direta (sem revisão)		Revisão antes ou depois da atualização			
	Notificação por email	Não		Sim (inscrição em lista)		Sim, personalizada			
Estruturação do conhecimento	Critério de estruturação	Temporal		Por assunto		Por autor		Outros (usuário etc)	
	Tipos de conteúdo	Resumos	Artigos	Dados	Problemas	Eventos	Multimídia	Trails outros	
	Direção dos apontadores	Unidirecional			Bi-direcional				
	Tipos de relacionamento	Não especificado		Generalização		Sumarização	Todo/parte		Outros
	Manutenção da consistência	Não disponível		Via procedimentos manuais		Checagens semi-automáticas		Checagens automáticas	
	Identificação do conhecimento ausente	Nenhum suporte			Listas de termos ausentes				
	Escopo do sistema	Fechado			Aberto				
Apresentação	Versionamento	Nenhum		Publicações ativas		Publicações "arquivadas"		Publicações "aposentadas"	
	Opções de visualização	Nenhuma		Mapa do sítio			Mapa de nós gráfico		
	Opções de navegação	Nenhuma		Hit List			Nós aleatórios		
	Mecanismos de Filtro	Nenhum	Contagem de acessos		Extração de nós baseada em ranking		Bubble-up	Outros	
	Internacionalização	Não	Interface multilingüe			Interface e conteúdo multilingüe			
	Personalização	Nenhuma	Interface do usuário		Mecanismos de filtro		Perfil do usuário armazenado		
	Separação do conteúdo da apresentação	Nenhuma	Apresentações múltiplas e não-lineares do mesmo conteúdo			Geração dinâmica de apresentações		Apresentações sensíveis ao contexto	
	Configuração cliente	Navegador Web		Navegador Web com JavaScript		Navegador com Java Applet		Navegador com plug-in	Navegador customizado
	Formato do documento	PDF		PostScript		HTML		Texto	Outros
Versão para impressão	Nenhuma		Issues			Volumes			
Gerenciamento	Seletividade	Inclusiva (disponível ao público)			Exclusiva (participação condicionada à qualificação ou perfil)				
	Autenticação	Nunca requisitada		Requisitada para algumas funcionalidades			Sempre		
	Mecanismos de autenticação	Nome do usuário e senha		Cookies		Certificados		Domínio	
	Contribuições anônimas	Não			Sim				
Tecnologia	Modelo de financiamento	Inscrição		Pagar por artigo		Pagar por submissão	Propaganda	Grants	
	Tipo de servidor	Servidor de HTML		Banco de dados		Máquina de busca	Máquina de workflow	Máquina de inferência	
	Tipo de busca	Por autor, título etc		Por palavra-chave		Por conceito	Não textual	Outras	
	Ranking dos resultados	Nenhum		Casamento aproximado		Preferência do usuário		Outros	
	Padrões	Java	JavaScript	XML		ASP	DHTML	Outros	
	Autenticação do documento	Não			Sim				
	Armazenamento	Permanente			Temporário				
Suporte a sessões	Não			Sim					

Apêndice B

Modelagem conceitual do ambiente CEJ

O objetivo deste apêndice é descrever a modelagem de dados conceitual do ambiente CEJ, na qual foi utilizado o diagrama de classes da UML e a língua inglesa nas descrições textuais – para garantir a internacionalização deste trabalho.

Modelagem geral do sistema

No nosso modelo orientado a objetos, uma revista é representada pela classe **Journal**. Uma instância do ambiente CEJ permite a definição de várias revistas simultaneamente, ou seja, a classe *Journal* poderá ser instanciada N vezes. Uma revista pode definir diversos *templates* de processos de publicação (classe **WFTptProcess**). Um usuário do ambiente CEJ (classe **User**) pode atuar com diversos papéis (classe **Role**) em uma determinada revista.

Um artigo (classe **Article**) é publicado em uma revista e tem um tipo (classe **ArticleType**), por exemplo, um *artigo original*, um “*survey*”, uma *revisão da literatura* etc. Documentos (classe **Document**) são associados a um determinado artigo, e também possuem um tipo (classe **DocumentType**), que pode ser uma *proposta de trabalho*, uma *pré-publicação* etc. Uma anotação (classe **Annotation**) tem um tipo (classe **AnnotationType**), é criada por um usuário e anota um documento ou responde a uma outra anotação - um dos dois. Finalmente, um conjunto de anotações (classe **AnnotationSet**) poderá ser criado em algumas fases de um *workflow* de publicação, reunindo as contribuições geradas naquela fase.

Existe no sistema o conceito de grupo (classe **Group**), que é criado por um usuário e associado a uma revista. A associação entre a classe *Group* e a classe *User* é de muitos para muitos, já que um usuário poderá fazer parte de mais de um grupo. Na definição dos processos de publicação, pode-se restringir a participação dos usuários aos membros de um determinado grupo de trabalho.

Um documento é composto por elementos (classe **DocElement**), que por sua vez são compostos opcionalmente por atributos (classe **DocAttribute**). Estas duas classes representam a estrutura de um documento XML no repositório de dados.

em uma determinada *revista* (classe **Journal-Role**). Por último, o sistema guarda um log (classe **Log**) da execução dos casos de uso do sistema por cada usuário.

Modelagem da máquina de *workflow*

- Definição de processos de publicação

A definição de um processo de publicação é realizada através da criação de um *template* de processo (classe **WFTptProcess**), associado a uma revista (classe **Journal**), e dos templates das tarefas que irão compor o processo (classe **WFTptTask**). Um template de tarefa é sempre associado ao conjunto de papéis de usuários (classe **Role**) que poderão executar a tarefa, e a um tipo de tarefa (classe **WFTaskType**), que é a entidade que implementa a tarefa. Por exemplo, o template para uma tarefa de "Submissão de proposta de trabalho" poderia ser associada a um tipo de tarefa "Submeter documento".

A um determinado tipo de tarefa poderão estar associados parâmetros de entrada e saída (classe **WFTaskParameter**), que poderão ser representados por qualquer objeto em Java: uma String, um inteiro, um documento (classe **Document**), uma conjunto de anotações (classe **AnnotationSet**) etc. A inclusão de um template de tarefa no template de um processo dispara a inclusão de um template de item de dados (classe **WFTptDataItem**) correspondente a cada parâmetro de saída definido pelo tipo da tarefa. A inclusão de um template de tarefa também dispara a associação desta tarefa com os templates dos itens de trabalho que farão parte da tarefa (classe **WFTptTaskItem**). Estes objetos representam uma associação entre o template do item de dados incluindo no workflow por ocasião da inclusão do template de tarefa, e o parâmetro correspondente definido pelo tipo da tarefa.

No ambiente CEJ, existem diversos tipos de tarefa pré-definidos: submeter documento, revisar documento, aprovar documento, comentar documento, aprovar documento, publicar documento etc. Ao definir uma tarefa em um determinado template de processo, pode-se associá-la a algum desses tipos de tarefa existentes, ou a algum novo tipo.

As tarefas de um *template* de processo podem ser visualizadas como nós de uma rede, que são interligados através de arestas ou *links* (classe **WFTptLink**). Diversos *links* poderão partir de uma única tarefa, de tal forma que o término de uma tarefa poderá disparar a execução de diversas outras tarefas. Por outro lado, uma tarefa só

poderá ser iniciada quando todas as tarefas que apontam para ela tiverem sido concluídas. Esta não é a única abordagem para o disparo de uma atividade, mas foi a escolhida no projeto do gerente de workflows do CEJ. Existe ainda um tipo de tarefa especial, pré-definido no ambiente CEJ e denominado “Tarefa condicional”, que permite avaliar um objeto do tipo booleano e direcionar o fluxo de trabalho para uma tarefa ou outra, de acordo com o valor do objeto - verdadeiro (*true*) ou falso (*false*). A figura a seguir ilustra o diagrama de classes do *workflow* do ambiente CEJ.

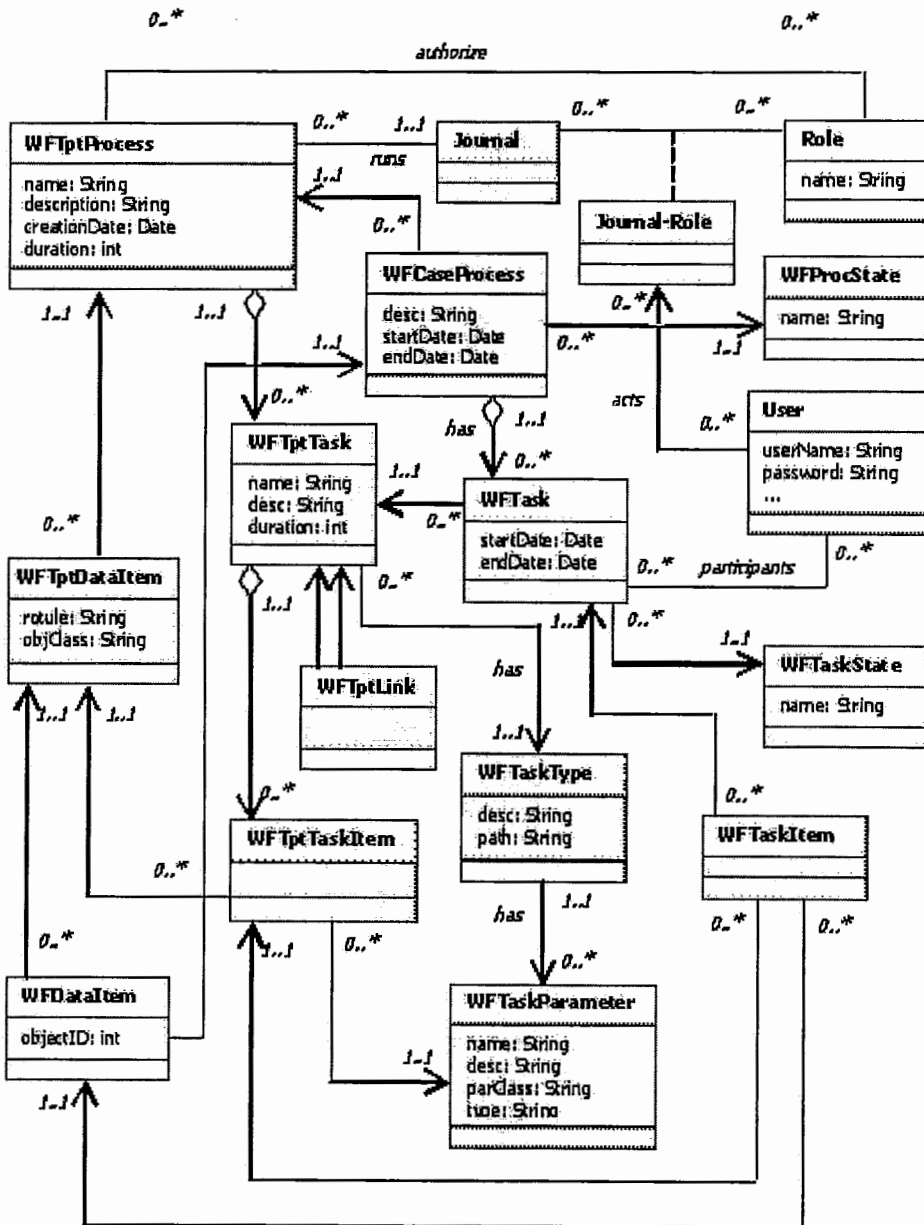


Figura B.2 – Modelo conceitual da máquina de *workflow* do CEJ

Assim como acontece com a definição das tarefas, existem também requisitos na adição de *links* entre as tarefas. Um link partindo de uma determinada tarefa X só poderá ser direcionado para uma tarefa Y se a tarefa Y tiver prioridade maior do que a da tarefa X, ou seja, se Y for cronologicamente mais velha.

- Execução de processos de publicação

A execução de um processo de publicação é realizada através da criação de um caso de processo (classe **WFCaseProcess**), que é associado a um template de processo (classe **WFTptProcess**) descrito anteriormente. Um caso de processo está sempre associado a um estado (classe **WFProcessState**), inicialmente o estado "aguardando". O acionamento do método *start()* de um caso de processo coloca o mesmo no estado "executando", disparando a criação das tarefas (classe **WFTask**) do template do processo, que são associadas aos templates de tarefa (classe **WFTptTask**).

As tarefas criadas são também associadas a um estado (classe **WFTaskState**), inicialmente no estado "aguardando". A execução de cada tarefa, assim como acontece com o caso de processo, depende do acionamento do método *start()* da tarefa. Uma tarefa é associada a um conjunto de usuários (classe **User**) que serão os executores da tarefa, os quais devem atuar com pelo menos um dos papéis definidos no template da tarefa. Finalmente, ao criar-se uma tarefa, automaticamente ela é associada aos itens de trabalho que farão parte da execução da tarefa (classe **WFTaskItem**). Cada item de trabalho que participa de uma tarefa está associado ao template do item de trabalho da tarefa (classe **WFTptTaskItem**) e à instância do template do item de dados (classe **WFDataItem**) que participa do caso do processo (classe **WFCaseProcess**).