

PROJETO SISTEMAS AVANÇADOS PARA COMUNICAÇÃO ELETRÔNICA – SOFTWARE ABERTO DE CORREIO, AGENDA E CATÁLOGO

Cláudio F. R. Geyer (geyer@inf.ufrgs.br) – II/UFRGS
Clairmont Borges (clermont@inf.ufrgs.br) – II/UFRGS
Patrícia Kayser Vargas (kayser@lasalle.tche.br) – UNILASALLE
Débora Nice Ferrari (nice@lasalle.tche.br) – UNILASALLE
Christiano Otero Ávila (chris@polaris.ucpel.tche.br) – UCPel
Adenauer Correa Yamin (adenauer@polaris.ucpel.tche.br) – UCPel
Cristiano Cachapuz e Lima (cristiano@urcamp.tche.br) – URCAMP
Luiz Cláudio Dalmolin (dalmolin@urcamp.tche.br) – URCAMP
Ricardo Balinski (balinski@procergs.br) – PROCERGS

RESUMO

Este artigo descreve os trabalhos que serão desenvolvidos no âmbito de um projeto de pesquisa aprovado no Edital 06/2000 da FAPERGS. Este projeto contempla uma parceria entre a PROCERGS e as seguintes universidades: UFRGS, UNILASALLE, UCPel e URCAMP. O objetivo do projeto é o aperfeiçoamento do sistema DIRETO, o qual oferece serviços eletrônicos de correio, agenda e catálogo. É esperado que os resultados do projeto contribuam na consolidação do DIRETO como um software de alta tecnologia, atendendo às necessidades da PROCERGS e de outras instituições/empresas que venham utilizá-lo. Considerando que sua distribuição será na forma de software livre, a estimativa é que o DIRETO irá atingir mais de 50 mil usuários somente no setor público estadual.

ABSTRACT

This article describes the works that will be developed in the scope of a research project approved in the Edictal 06/2000 of FAPERGS. This project contemplates a partnership among PROCERGS and the following universities: UFRGS, UNILASALLE, UCPel and URCAMP. The objective of the project is the improvement of the DIRECT system, which offers electronic services of mail, calendar and catalog. It's expected that the results of the project contribute in the consolidation of the DIRECT as a software of high technology, assisting to the needs of PROCERGS and other institutions/enterprises that may use it. Considering that its distribution will be in the form of free software, the estimate is that the DIRETO will reach more than 50 thousand users only in the state public sector.

1 INTRODUÇÃO

O uso de redes de computadores, e da Internet de forma especial, potencializa a disponibilização de informações e estimula mudanças nos processos das mais variadas áreas do conhecimento. Para que essa interação ocorra são utilizados sistemas especializados para suporte à troca de informações eletrônicas e/ou para realização de transações computacionais de forma distribuída. Normalmente, as redes de computadores apresentam características heterogêneas. Essa heterogeneidade ocorre tanto em relação ao hardware quanto ao software, sendo essa situação válida não somente para os ambientes acadêmicos como para os comerciais. Em grandes empresas essa heterogeneidade costuma ser mais complexa quando existe um histórico de vários anos de uso da informática, e onde ainda são mantidos sistemas para *mainframe* em funcionamento.

Em teoria, essa diversidade de componentes possibilitaria um maior número de configurações, cada uma das quais mais adequada ao desempenho de uma determinada tarefa ou para uma determinada situação. Porém, na prática nota-se que o controle dessas diferentes configurações de componentes é bastante complexo. A computação em ambiente distribuído é complexa, sujeita à falhas e cara para desenvolver, testar e manter. Para minimizar isto, surgiram diversas alternativas para simplificar a modelagem e a implementação em sistemas distribuídos heterogêneos ([10] e [15]).

Deste modo, uma das preocupações atuais, tanto em ambientes acadêmicos como empresariais, consiste na disponibilização de informações distribuídas (o que pode ser obtido através de replicação dos dados e do seu acesso através de transações distribuídas) bem como a necessidade de padronizar a forma de comunicação entre os membros dos grupos de trabalho. Para padronizar a comunicação, deve-se obter um produto viável em

termos de custos econômicos ao mesmo tempo com características refinadas, como filtros adaptáveis aos perfis de usuário que facilitem o uso.

Outra motivação importante para esse trabalho é o contexto particular da PROCERGS. A PROCERGS e a maioria de seus clientes do Estado trabalham desde 1992 com ferramentas proprietárias de correio eletrônico. Isso causou problemas tais como: (a) dificuldades para comunicação interna e para implementação de ferramentas de *workflow*; (b) heterogeneidade dificulta o treinamento dos usuários; (c) custo elevado; (d) falta de integração entre os componentes dos sistemas.

A PROCERGS então decidiu desenvolver sua própria solução materializada através do produto DIRETO, que se baseou totalmente em protocolos padrão Internet. Além disso, o DIRETO será um software livre, com distribuição e licenças gratuitas, a partir de um site a ser construído e mantido pela PROCERGS, segundo os termos da GNU *General Public License* (GPL). O DIRETO está em fase de desenvolvimento, possuindo os recursos básicos e necessários para garantir o seu funcionamento. Porém, recursos mais avançados, os quais servem de motivação para o presente projeto, serão implementados em versões futuras do produto através dessa cooperação.

2 APRIMORAMENTOS NO DIRETO

A PROCERGS realizou um levantamento das necessidades que o produto deveria suprir. Estas necessidades foram divididas em gerais e de ambiente e a partir delas o sistema iniciou seu desenvolvimento. O DIRETO é um produto modularizado que utiliza os protocolos *vCalendar* para agenda, *SMTP* e *IMAP* para correio e *LDAP* para catálogo. Ele está sendo desenvolvido utilizando a linguagem de programação Java e também emprega outros programas de distribuição livre, como os sistemas operacionais *GNU Linux* e *FreeBSD*, o banco de dados *PostgreSQL*, entre outros.

A primeira versão está em uso pelos funcionários da PROCERGS e novas necessidades estão sendo detectadas. Assim, o objetivo deste projeto é fornecer subsídios para o desenvolvimento e a incorporação das seguintes características ao produto DIRETO: (a) Documentação utilizando UML; (b) Metodologia para testes; (c) Avaliação e otimização de desempenho; (d) Filtros adaptáveis aos perfis de usuários; e (e) Transações distribuídas. Esses objetivos irão contribuir para a produção de um sistema mais eficiente e tecnologicamente mais bem acabado. Cada um desses objetivos será detalhado nas seções seguintes.

2.1 Documentação Utilizando UML

Para facilitar a manutenção e agilizar o processo de inclusão de novas características no DIRETO é fundamental a realização de atividades relacionadas à documentação e testes. Por isso, uma das primeiras tarefas envolve a realização da documentação. Em um primeiro momento houve uma interação entre a equipe da PROCERGS e alguns representantes das instituições a fim de obter uma visão geral do sistema. A partir disso os códigos foram submetidos a uma ferramenta de engenharia reversa a fim de obter um diagrama de classes em notação UML (*Unified Modeling Language*) [7]. Esta é uma linguagem para especificação e modelagem de sistemas orientados a objetos. As etapas seguintes envolvem a análise do código a fim de obter outras visões do sistema, como por exemplo através de diagramas de seqüência. Também estão sendo desenvolvidos diversos textos HTML descrevendo tanto aspectos de instalação quanto de codificação.

2.2 Metodologia para Testes

Para a realização de testes será desenvolvida uma metodologia apropriada. A partir da especificação em UML obtida na fase de documentação serão definidos subsídios para a construção da metodologia. Essa fase permitirá validar e avaliar diferentes aspectos do sistema. Também é um dos objetivos dessa tarefa consolidar as propostas de material de treinamento oriundas dos diversos grupos que integram o projeto à medida que o mesmo se desenvolve.

2.3 Avaliação e Otimização de Desempenho

Para motivar o uso do DIRETO é necessário otimizar o sistema com relação ao desempenho. Por se tratar de um sistema complexo, somente uma prática de avaliações sistemáticas pode conduzir a otimizações efetivas [8]. Por exemplo, um ponto a ser avaliado é a eficiência do atual *parser* de XML.

2.4 Filtros Adaptáveis aos Perfis de Usuários

Atualmente, o excesso de informação é um problema comum para usuários de redes de computadores. Seja usando correio eletrônico, acessando servidores de notícias (*newsgroups*) ou servidores WWW, os usuários

destes serviços lidam constantemente com o problema de encontrar ou selecionar informações relevantes. Técnicas de Filtragem (ou Classificação) em [9], [13] e [14] são bastante usadas para organizar as informações ([5], [6] e [3]), visando ajudar usuários a identificarem mais facilmente as informações que procuram ou desejam. Desta forma, a inclusão de filtros adaptáveis aos perfis de usuários no DIRETO tem os seguintes objetivos: (a) permitir que o usuário especifique seu perfil inicial de preferências (quanto ao conteúdo que deseja filtrar na caixa de entrada do correio eletrônico); (b) realizar a filtragem para cada usuário, baseando-se no seu perfil de preferência e também considerando a origem e o conteúdo do correio eletrônico; e (c) adaptar continuamente o serviço de filtragem, usando como feedback as informações oriundas dos perfis de usuários.

O serviço de monitoramento e construção dos perfis de usuários e grupos de usuários é outro ponto importante [6]. Devido à incerteza existente na construção de perfis, este projeto tem como um de seus objetivos pesquisar as abordagens estatísticas e probabilísticas usadas na construção de perfis de usuários, assim como os formalismos usados para modelar e manipular incerteza ([16] e [1]). Esta pesquisa visa aperfeiçoar e dar suporte ao serviço de filtragem de correio eletrônico, construindo um perfil para cada usuário do software DIRETO, e criando perfis de grupos de usuários, através da similaridade de seus interesses e comportamentos. Uma vez criados estes perfis de grupos de usuários, serão aplicadas as técnicas de filtragem colaborativa para aumentar a sinergia destes grupos ([12], [11], [2], [6] e [17]).

2.5 Transações Distribuídas

No contexto de uma comunicação distribuída, e cliente-servidor em especial, raramente as requisições são constituídas de operações únicas ou várias requisições executadas como operações independentes, bem como normalmente os servidores têm a necessidade de atender diversos clientes simultaneamente. Por isso, o serviço de transações é oferecido em modelos cliente-servidor com o objetivo de permitir consistência, durabilidade, isolamento e integridade dos dados (propriedades ACID). O uso de transações em sistemas distribuídos serve para endereçar dois problemas: concorrência e falhas ([4] e [18]). O DIRETO, por ser um sistema distribuído, também possui a necessidade de realizar um tratamento de transações. Desta forma, este projeto visa o estudo e implementação de técnicas de transações distribuídas, em especial na organização de reuniões através do serviço de agenda, sendo o controle destes aspectos vital para manter a consistência do sistema.

3 CONCLUSÃO

O projeto como um todo se propõe a melhorar a comunicação e a divulgação de informações entre os diversos órgãos públicos estaduais além de poder ser usado também em outros contextos usuários – órgãos públicos em geral, universidades, escolas, etc. – por se tratar de um software livre.

A tecnologia desenvolvida pela PROCERGS, e aquela aplicada no DIRETO em especial, serão repassadas para as universidades e portanto para os alunos de um modo geral contribuindo para uma visão mais aplicada dos diversos aspectos estudados academicamente.

Por sua vez, a tecnologia desenvolvida pelas universidades através dos seus núcleos de pesquisa será repassada à PROCERGS que poderá decidir pela sua integração ou não ao seu processo de desenvolvimento.

É importante salientar que além do DIRETO, os subprodutos tecnológicos gerados tanto pela PROCERGS quanto pelas universidades serão disponibilizados para a comunidade em geral, uma vez que o projeto será conduzido seguindo as premissas que regem o desenvolvimento de sistemas categorizados como software livre.

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] Machine Learning. **In:** American Association of Artificial Intelligence. 2000. Disponível em <http://www.aaai.org/Pathfinder/html/machine_learning.html>. Acesso em 12 dez. 2000.
- [2] BILLSUS, Daniel; PAZZANI, Michael J. Learning Collaborative Information Filters. Int. Conf. on Machine Learning. **Proceedings...** Morgan Kaufman, Madison, Wiscosin. 1998. Disponível em <<http://www.ics.uci.edu/~pazzani/Publications/MLC98.pdf>>. Acesso em 20 jan. 2000.
- [3] BRADLEY, P. S., SARAWAGI, S., FAYYAD, U. M. (Ed.). Internet Data Mining. **In:** SIGKDD Explorations – Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, June 2000. Volume 2, Issue 1. Disponível em <<http://www.acm.org/sigs/sigkdd/explorations/issue2-1/>>. Acesso em 15 fev. 2001.
- [4] CAMPOS, J. B. de A. **Um Servidor de Arquivos Distribuído com Suporte a Transações**. CPGCC/UFRGS, 1996. 141p. Dissertação de Mestrado.

- [5] FAORO, M.A., GEYER, C.F.R. **Um Agente Para Localização e Organização de Informações**. V Semana Acadêmica do PPGC, Instituto de Informática/UFRGS. Disponível em <<http://www.inf.ufrgs.br/pos/SemanaAcademica/Programa-SA2000.html>>. Acesso em: 29 abr. 2001.
- [6] HORVITZ, E. Agents with Beliefs: Reflections on Bayesian Methods for User Modeling, pp.441. **In:** (Editors) Jameson, A.; Paris, C. and Tasso, C. User Modeling – Proceedings of the Sixth International Conference (UM97) Springer-Verlag, 1997.
- [7] JACOBSON, I., BOOCH, G., RUMBAUGH, J. **The Unified Software Development Process**. Addison-Wesley, 1999.
- [8] Killelea, P. **Web Performance Tuning**. Sebastopol: O'Reilly & Associates, 1998. 368p.
- [9] NIGAM, K. et al. Text Classification from Labeled and Unlabeled Documents using EM. **In:** Machine Learning Journal, 1-34. Kluwer Academic Publishers. 1999. Disponível em <<http://www.cs.cmu.edu/~knigam/papers/emcat-mlj99.ps>>. Acesso em: 31 jan. 2000.
- [10] OBJECT MANAGEMENT GROUP. **CORBA Definitions**. Disponível em <<http://www.omg.org>>. Acesso em: 01 mar. 2000.
- [11] PAZZANI, Michael J.. A Framework for Collaborative, Content-Based and Demographic Filtering. **In:** Artificial Intelligence Review. 1999. Disponível em <<http://www.ics.uci.edu/~pazzani/Publications/AIREVIEW.pdf>>. Acesso em: 20 jan. 2000.
- [12] PENNOCK, D.M., HORVITZ, E. Collaborative Filtering by Personality Diagnosis: A Hybrid Memory- and Model-Based Approach. **In:** IJCAI Workshop on Machine Learning for Information Filtering, International Joint Conference on Artificial Intelligence (IJCAI-99), August 1999, Stockholm, Sweden. Disponível em <<ftp://ftp.research.microsoft.com/pub/ejh/cfpd.pdf>>. Acesso em: 29 abr. 2001.
- [13] SAHAMI, M et al. **A Bayesian Approach to Filtering Junk Email**. AAAI'98 Workshop on Learning for Text Categorization, July 27, 1998, Madison, Wisconsin. Disponível em <<http://www.research.microsoft.com/~sdumais/spam98.ps>> Acesso em: 26 jan. 2000.
- [14] SEGAL, R. B., KEPHART, Jeffrey O. MailCat: an intelligent assistant for organizing e-mail. **In:** Proc. Third Annual Conference on Autonomous Agents, May 1-5, 1999, Seattle, WA USA, p.276-282.
- [15] SUN Microsystems. **The Source for Java technology**. Disponível em <<http://java.sun.com/>> . Acesso em: 03 mar. 2001.
- [16] Uncertainty in Artificial Intelligence. **In:** Proceedings of the 15th UAI Conference. 1999. Disponível em: <<http://www.auai.org/>>. Acesso em: 29 abr. 2001.
- [17] User Modeling. **In:** Proceedings of the Seventh International Conference. Ed. Judy Kay, Springer Wien, New York, 1999. Disponível em: <http://www.um.org/um_99/Proc/>. Acesso em: 29 abr. 2001.
- [18] WEIHL, W. E. Transaction-Processing Techniques. **In:** Distributed Systems, 2nd ed., Sape Mullender (ed). Addison-Wesley, 1993. p.329-352.