

Canais de Informação com Serviços de Filtragem no Direto

Ricardo Balinski¹, Clairmont Borges², Cláudio F.R. Geyer²

¹PROCERGS – Companhia de Processamento de Dados do Estado RS
Praça dos Açorianos S/Nº – 90010-340 – Porto Alegre – RS – Brazil

²Instituto de Informática – Universidade Federal do Rio Grande do Sul (UFRGS)
Caixa Postal 15.064 – 91.501-970 – Porto Alegre – RS – Brazil
{clermont,geyer}@inf.ufrgs.br, balinski@via-rs.net

Abstract. *This paper describes the development of Information Channels with Filtering Services (ICFS) in the software Direto. Once Direto will be used in most government services in the state of Rio Grande do Sul, the ICFS provides usefull information to the population. The ICFS service is being developed as a MsC. Dissertation in the join research project between PROCERGS and Informática-UFRGS.*

Resumo. *Este artigo descreve o desenvolvimento de Canais de Informação com Serviços de Filtragem (CISF) no software Direto. Uma vez que o Direto será usado na maioria dos serviços de governo no Estado do Rio Grande do Sul, o CISF provê informações úteis para a população. O serviço de CISF está sendo desenvolvido como uma Dissertação de Mestrado no projeto de pesquisa conjunto entre PROCERGS e Informática-UFRGS.*

1. Introdução

Vivemos numa sociedade onde somos educados para acreditar e seguir os importantes valores da democracia, da liberdade de expressão, da livre iniciativa e da livre competição. Para colocar estes princípios fundamentais em prática, é preciso criar sistemas de informação que atendam às Liberdades de Expressão e Comunicação [BORGES 2000] da sociedade. É preciso aproximar o estado e as forças produtivas, criando maior transparência na aplicação dos impostos recolhidos e aumentando a possibilidade de fiscalização por parte da sociedade civil organizada e comprometida com o progresso e a justiça social.

Atualmente, o mercado é muito dinâmico e as tomadas de decisão, assim como os fluxos de informação, são processos quase contínuos. Tudo evolui, incluindo os novos serviços jornalísticos baseados na Internet – os quais atendem nichos e necessidades cada vez mais específicos, ao mesmo tempo em que abrem novas frentes de empreendimento e trabalho para as novas gerações de jornalistas.

Por outro lado, o software está presente em praticamente todas as atividades e relações humanas. O software, além de ser um conjunto de instruções de computador, é também uma representação de conhecimento – e conhecimento é Poder. Os estados democráticos estão fundamentados na contínua busca do equilíbrio entre os principais poderes. Sendo o software uma nova forma de poder, o Software Livre permite promover o crescimento econômico com justiça social [BORGES 2001], juntamente com estes serviços jornalísticos de nova geração.

Visando implementar estas estratégias, foi desenvolvido o software Direto [DIRETO 2000] pelo Governo do Estado do Rio Grande do Sul. Logo depois, se alinhando com estas diretrizes progressistas, surgiu o projeto conjunto "Projeto Direto" entre PROCERGS e Informática-UFRGS [PROJETO 2001, GEYER 2001]. Conseqüentemente, uma das principais frentes de trabalho neste projeto conjunto são os Canais de Informação com Serviços de Filtragem (CISF), em desenvolvimento no âmbito de um trabalho de dissertação de mestrado. A seguir são apresentados trechos representativos do trabalho do mestrando Ricardo Balinski.

2. A Arquitetura do Serviço

Os serviços públicos são grandes produtores de informação e estão distribuídos geograficamente, embora estejam organizados hierarquicamente. Por outro lado, os consumidores de informação (empreendedores, assalariados, profissionais liberais, etc.) possuem demandas diferentes, seja no tipo do conteúdo, ou seja no grau de especificidade deste conteúdo. Neste cenário, existe ainda a complexidade do fato dos consumidores de informação também estarem geograficamente distribuídos, e o fato das fontes produzirem continuamente. Este problema foi resolvido no trabalho [BORGES 1997] e serve de base para o desenvolvimento dos Canais de Informação com Serviços de Filtragem (CISF) no Direto. Nas subseções seguintes este serviço é descrito sucintamente.

2.1 Produção de Informação

Cada vez que um serviço de governo produz uma informação útil e de interesse público, esta informação é encaminhada de forma automática (ou assistida) para um ou mais Canais de Informação – conforme a origem da fonte e também conforme o conteúdo da mensagem. Por exemplo, se uma informação sobre campanha de vacinação é produzida pela Secretaria da Saúde, ela é encaminhada para o canal Saúde Pública, a partir daí é disseminada pela rede.

2.2 Consumo de Informação

Quanto ao consumo, o usuário acessa o Serviço de Canais através de um portal na Web, para se cadastrar em um ou mais Canais de seu interesse. A partir daí ele recebe via correio eletrônico as informações disseminadas nos Canais em que se cadastrou. Opcionalmente, para cada Canal, o usuário do serviço pode criar Filtros, para poder focalizar sua atenção em informações mais específicas. Por exemplo, uma empresa jornalística de nova geração que produz análises econômicas sobre o mercado de criação de gado de corte, pode criar Filtros para este foco de interesse nos respectivos Canais em que se cadastrou.

2.3 Viabilidade do Serviço

A diferença entre os serviços CISF e as listas comuns de distribuição de correio eletrônico é fundamentalmente a viabilidade. A arquitetura que baseia os serviços CISF permite crescimento contínuo e com custo incremental nas operações de Coleta, Filtragem e Disseminação de informação, mesmo em cenários complexos conforme o descrito no início desta seção 2.

3. Os Serviços CISF no Direto

Para não fugir do escopo (de esforço de uma dissertação de mestrado) foram realizadas algumas simplificações para testar algumas funcionalidades usando as tecnologias que atualmente sustentam o Direto.

3.1 Fluxo do Serviço

Quando uma informação é produzida, a mesma é enviada ao moderador do canal, caso o canal seja moderado. O moderador avalia o texto e, caso seja aprovado, define algum horário para envio aos assinantes. O texto é enviado ao endereço do canal que vai redirecionar para todos seus assinantes. Quando o texto chega à caixa postal do assinante, é realizada a comparação entre o seu conteúdo com o filtro do usuário. Se o texto possuir uma similaridade maior ou igual a desejada pelo usuário, o mesmo é entregue para o usuário, caso contrário o texto é descartado da caixa postal do usuário.

3.2 Critério de Filtragem

Primeiramente é feito um pré-processamento no texto das informações, removendo termos irrelevantes, transformando os termos em radicais da língua portuguesa, e substituindo-os por sinônimos mais representativos. Após estas transformações, o termo resultante é inserido em um vetor, juntamente com o seu peso. O método utilizado para definir o peso é o método *word frequency weighting*, o qual associa o peso do termo com a sua frequência (ou importância) no texto.

Posteriormente, é comparada a similaridade entre o vetor que representa a informação com o vetor que representa o filtro do usuário do canal. A comparação é feita utilizando o método que calcula o cosseno do ângulo formado pelos vetores. Neste método, a informação é representada por termos e seus respectivos pesos, mapeados num espaço vetorial, onde cada dimensão é definida por um termo. Uma informação é filtrada para um usuário de canal quando o vetor da informação é similar ao vetor do filtro, isto é, estão próximos no espaço vetorial – cuja medida é obtida pelo cosseno do ângulo formado entre eles.

3.3 Considerações Tecnológicas

O serviço de canais de informações está totalmente baseado na infra-estrutura do Direto, e utiliza os seus serviços de *LDAP*, *SMTP* e *IMAP* para distribuir as informações. Todo o processamento e os dados ficam localizados no servidor. Isso permite o usuário usar o serviço independentemente do cliente de correio ou do navegador *Web*, pois todos estes serviços se baseiam em protocolos padronizados. A Figura 1 ilustra o serviço de Filtragem no Direto.