

Geração Dinâmica de Conteúdo Personalizado para Sites de Notícias

Jonh Ney Barbosa de Sousa, Paulo Artur de Sousa Duarte, Gisele Pereira dos Santos, Ítalo Romeiro Wanderley, Francisco Ari Josino Júnior, Priscylla da Silva Tavares, Davi Jucimon Monteiro, Cidley Teixeira de Souza

IFCE – Instituto Federal de Educação, Ciência e Tecnologia do Ceará
Av. 13 de Maio, 2081 – Benfica – Fortaleza/CE – Brasil

+55 85 88158553

E-mail: {johnnbs88, neopaulossj, gijeleps, romeirao, ririari, priscyllatavares, jucimon, cidley}@gmail.com

ABSTRACT

The press newspaper is suffering a severe crisis that was established with the advent of the Internet, which brought the spread of free high quality news. Even their online versions are facing major problems to gather resources. Experts believe that the future of online newspapers is a greater interaction with the user and the use of personal advertising. This paper presents an architectural model for generating personalized content oriented for news sites. Therefore, this aims at increasing the flexibility of the structure of online newspapers in order to encourage all sides of this issue, increasing convenience to the user while generating a more robust revenue stream to the newspapers.

Categories and Subject Descriptors

H.5.3 [Information Interfaces and Presentation]: Group and Organization Interfaces – *Web-based interaction*.

General Terms

Management, Design, Economics, Standardization.

Keywords

Newspaper On-line, News Sites, Personalization, Web 2.0

RESUMO

O jornal impresso sofre com uma forte crise que se estabeleceu com o advento da Internet, que trouxe a propagação de notícias de qualidade de forma gratuita. Mesmo as suas versões on-line enfrentam grandes problemas para captarem recursos. Especialistas acreditam que o futuro dos jornais on-line passa por uma maior interatividade com o usuário e com o uso de publicidade personalizada. Este artigo pretende apresentar um modelo de arquitetura de geração de conteúdo personalizado voltado para sites de notícia. Com isso, este modelo pretende aumentar a flexibilidade da estrutura dos jornais on-line de modo a favorecer todos os lados desta questão, aumentando a comodidade para o usuário ao mesmo tempo em que gera uma fonte de receita mais sólida para os jornais. O artigo mostrará as funcionalidades e facilidades desse modelo de

arquitetura e apresentará sua aplicação e impacto, uma vez adotado o modelo em estudo.

1. INTRODUÇÃO

Com o advento da Internet, as pessoas mudaram os seus hábitos. Passou a ser vital estar constantemente atualizado, de forma que as pessoas passaram a buscar informações várias vezes ao dia. Isso faz com que, muitas vezes, as versões impressas dos jornais publiquem notícias que, já totalmente esmiuçadas por outras mídias, não atraíam tanto a atenção do leitor, mais interessado nos acontecimentos do dia atual que do anterior. Esta é uma das causas da grave crise pela qual está passando o mundo do jornal impresso, com a crescente queda nas vendas dos jornais, tanto no exterior quanto no Brasil [1].

A solução encontrada pelos jornais foi adaptar-se aos novos tempos, criando portais de notícias, de forma a não perder o público que migrou para a Internet e conseguir o dinheiro através de anúncios de publicidade e com o estímulo a assinaturas, o qual seria incentivado pelo conteúdo extra que seria fornecido para os assinantes.

No entanto, essa fórmula mostrou-se deficiente: o conteúdo pago afastava visitantes, diminuindo tanto a relevância do site quanto o dinheiro ganho através de publicidade. Por sua vez, o conteúdo gratuito carecia de fontes de receita, uma vez que a verba de anunciantes era insuficiente para suprir a perda de receita com o declínio das vendas do jornal impresso. Além disso, os sites de notícias também passaram a sofrer com a concorrência dos blogs, que, por geralmente serem voltados a nichos específicos, oferecem um conteúdo mais direcionado ao seu público-alvo, e das redes sociais, que oferecem um dinamismo maior e, muitas vezes, com notícias mais atualizadas. Muitas foram as maneiras que os jornais tentaram para solucionar esse problema [2], no entanto, nenhuma firmou-se como um modelo predominante.

Uma idéia lançada pelo diretor geral da Google, Eric Schmidt, mostra bem o futuro dos jornais on-line. Segundo ele, são vários os modos de se manter a sustentabilidade em sites de notícias. Contudo, o modelo que predominará será o de conteúdo gratuito sustentado por publicidade [3]. Em sua visão, os jornais poderiam criar novas formas de produção de sites, tendo como objetivo proporcionar uma maior interatividade entre o usuário e o site de notícias. Seguindo essa linha de pensamento, os anúncios de publicidade também seriam veiculados de forma personalizada, contextualizados com os temas das notícias e as preferências do leitor, de forma a aumentar a probabilidade do usuário se interessar pelo produto anunciado.

Esse é o objetivo do estudo, oferecer opções de construção de sites de notícias que proporcionem ao usuário uma maior facilidade de acesso às informações, e ao jornal uma estratégia de geração de renda que seja favorável ao progresso de seu empreendimento.

O restante deste artigo está organizado da seguinte maneira: a seção 2 dissertará sobre os referenciais teóricos que servem como base para o artigo. A seção 3 introduzirá os trabalhos relacionados, mostrando algumas soluções já existentes e suas deficiências. Na seção 4, haverá uma discussão sobre o ambiente de personalização proposto, apresentando a arquitetura proposta no estudo, detalhando seu funcionamento em um exemplo de uso e avaliando os progressos de sua implementação. Por fim, na seção 5 estarão as considerações finais, dissertando um pouco sobre as vantagens do uso da arquitetura proposta.

2. REFERENCIAL TEÓRICO

2.1 Personalização x Customização

Embora geralmente tratados como sinônimos no dia-a-dia, diversos autores pontuam que os termos personalização e customização têm significados distintos quando se trata das formas como um conteúdo e a aparência de um site são adaptados às preferências de cada usuário [4].

Na customização, o usuário é o agente modificador da página, pois lhe é dada a possibilidade explícita de controlar a aparência da página, alterando ou removendo elementos, priorizando ou preterindo determinados assuntos, configurando o que receberá de acordo com os seus gostos.

Já na personalização, o próprio site é o agente modificador da página, com o usuário contribuindo para os ajustes apenas implicitamente. Nela, o site monitora o comportamento do usuário enquanto este navega em suas páginas e, com base nesses dados, adiciona opções e elementos que aparentem estar de acordo com o perfil obtido deste usuário em questão.

Na arquitetura apresentada neste trabalho, optou-se por um modelo híbrido combinando personalização e customização. Ao mesmo tempo em que um dos objetivos principais é fazer com que o usuário tenha que esforçar-se ao mínimo para que o site fique de acordo com seus gostos, também lhe é dada a possibilidade de configurar quaisquer elementos que queira.

2.2 Ontologia

Ontologias são uma forma de representar conceitos de um determinado domínio, tendo como componentes: Classes, Indivíduos (objetos gerados a partir das Classes), Atributos e relacionamentos. Os relacionamentos em ontologias ocorrem quando indivíduos de uma determinada Classe são usados como Atributos para indivíduos de outras Classes. [5]

No sistema de geração de conteúdo, a utilização dessa metodologia possibilitará uma maior agilidade no gerenciamento dos conteúdos, já que utilizando-a será possível termos o tratamento de sub-classes de conteúdos (Esportes – Futebol) e o relacionamento entre essas classes com o objetivo de interligar os assuntos relacionados a determinada notícia.

3. TRABALHOS RELACIONADOS

A idéia de se colocar, em notícias e artigos, tags com palavras-chave relacionadas ao conteúdo destas já é algo bastante difundido na Internet, sendo utilizada por blogs e sites de notícias. No entanto, na maioria das vezes, essas tags pecam por serem bastante genéricas e por, às vezes, não seguirem critérios bem-definidos.

O Google tem vários serviços voltados à personalização e customização do conteúdo, tais como o Google News, que mostra as principais manchetes dos sites de notícias e possibilita ao usuário editar cada seção da forma que bem entender; o iGoogle, que oferece a possibilidade de customizar as posições em que o conteúdo da página principal aparecem com um simples arrasto do mouse; o GoogleNewsTimelines, que é um serviço que mostra o histórico de um determinado assunto na forma de uma linha de tempo; além do Gmail, que permite mudar deste o tema de fundo até as pastas onde os e-mails são armazenados.

4. AMBIENTE DE PERSONALIZAÇÃO PROPOSTO

O ambiente a ser personalizado é o site de notícias, no qual o usuário, devidamente cadastrado e logado, receberá uma página com os temas e assuntos que lhe são mais relevantes. A personalização acontecerá à medida em que o usuário acessar o site, como será explicado a seguir. Quais serão e como ocorrerão essas mudanças, serão definidas pela arquitetura de geração de conteúdo.

4.1 Arquitetura

Foi desenvolvida uma arquitetura como proposta de solução ao problema da personalização das notícias e publicidades. Essa arquitetura está dividida em módulos, como é mostrada na figura 1.

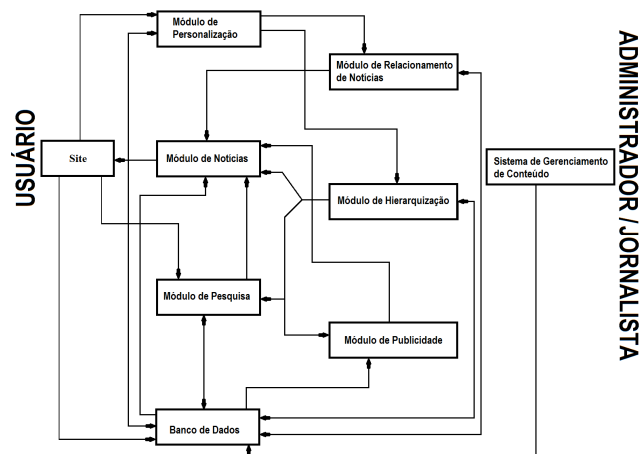


Figura 1 – Arquitetura

Para esta arquitetura, optou-se por um modelo misto entre personalização e customização, que, embora permita a customização de suas preferências, privilegie a personalização, realizando uma coleta indireta de informações sobre o usuário, enquanto este navega pelo site.

O cadastro de um novo usuário deve ser feito da maneira mais objetiva possível, perguntando apenas as informações mais básicas, tais como nome, e-mail, login e senha, evitando a possibilidade de um potencial usuário optar por não se cadastrar.

Ao acessar pela primeira vez, o usuário encontrará, inicialmente, em seu perfil uma página principal padronizada a todos os usuários. Esse perfil nada mais é que o site personalizado para o usuário em questão e é através dele que o sistema coletará as informações necessárias para a personalização.

A coleta de dados é feita pelo módulo de personalização. Assim que o usuário é logado, o módulo de personalização carrega as informações do usuário que estão no banco de dados, incluindo o

atual ranqueamento dos temas preferidos. Caso seja a primeira vez que acessa o sistema, os campos equivalentes a tais dados estarão em branco. A função deste módulo é monitorar a atividade do usuário, registrando cada acesso dele e atualizando os dados, que serão devidamente ranqueados no módulo de hierarquização. Ele também será encarregado de armazenar junto ao perfil do usuário as tags que este use com mais frequência para buscas mais específicas. Quando o usuário solicitar a sua saída do sistema, antes deste ser efetuado, o módulo atualiza o registro do usuário no banco de dados com os dados mais recentes presentes nele.

Nesta arquitetura, a customização também fica a cargo do módulo de personalização. O usuário poderá manipular a interface da página para um formato que lhe seja mais aprazível e, até mesmo, excluir ou incluir elementos de acordo com a sua vontade, aumentando a interatividade deste com o conteúdo que recebe.

O módulo de hierarquização é o grande responsável pela personalização do sistema. Ele recebe os dados atualizados do módulo de personalização e ranqueia, de acordo com a frequência de acessos, as categorias seguindo uma árvore de hierarquia pré-estabelecida. Essa árvore de hierarquia permite que cada notícia seja catalogada desde o tema mais geral até o sub-tema mais específico que a notícia aborda. Tal catálogo estaria em um campo do registro da notícia no banco de dados. Sempre que uma notícia é acessada, o contador da frequência de acessos do tema e sub-temas é incrementado. A classificação e ranqueamento das notícias deve ser feito de modo a evitar enganos que venham a favorecer um tema que não seja da real preferência do usuário.

O módulo de pesquisa é encarregado por personalizar as pesquisas feitas pelo usuário. Inicialmente, ele faz uma pesquisa ao banco de dados para recolher todas as notícias relacionadas à requisição do usuário. Então, ele, utilizando as informações do ranqueamento das preferências do usuário, obtidas no módulo de hierarquização, ordena as notícias de acordo com sua relevância, dando prioridade àquelas que sejam dos temas preferenciais do usuário, retornando ao módulo de notícias os resultados que se encontrem o mais próximo possível aos interesses do usuário.

O módulo de relacionamento de notícias será utilizado pelo provedor de conteúdos para estabelecer um elo entre notícias que demonstrem a evolução de um determinado acontecimento. O objetivo dele é estabelecer uma linha de tempo na qual o usuário pode visualizar o andamento de um certo evento no decorrer do tempo até aquele momento e possibilitar ao usuário acompanhar esses eventos sem a necessidade de ter que procurar por informações novas. Uma vez solicitado pelo perfil do usuário e registrado no módulo de personalização o acompanhamento de certo evento, sempre que uma notícia relacionada a esse assunto entrar no sistema, essa notícia “procurará” o usuário e não o contrário, evitando assim que ele não consiga acompanhar eventos de seu interesse por não conseguir encontrar as notícias sobre este.

O módulo de publicidade é o responsável pela geração de publicidade personalizada. Ela recebe os dados atualizados vindo do módulo de personalização e, então, faz uma triagem no banco de dados em busca dos anúncios publicitários que tenham ligação com as preferências do usuário, ranqueando-os, não apenas de acordo com as categorias de notícias, mas também pelo índice de acessos que tal anúncio teve com outros usuários e mesmo com o usuário em ação no momento.

O módulo de notícias é a interface entre toda essa arquitetura e o site de notícias. Ele fará a organização dos resultados retornados pelos demais módulos para exibição na página do perfil. Ele recebe

como parâmetros as categorias devidamente hierarquizadas e as notícias e as posições relativas delas no site, bem como o resultado retornado pelos módulos de relacionamento de notícias, publicidade e as atualizações devido à customização, e as organiza e envia para a exibição no perfil do usuário, encaixando todas as informações no layout do site. Assim, o módulo de notícias montará a página personalizada.

O objetivo dessa arquitetura é proporcionar ao usuário a antecipação de alguns passos na busca das notícias de seu interesse, aumentando a satisfação, fidelidade e o tempo de acesso do usuário a esse determinado site de notícias.

4.2 Exemplo de Uso

Segue abaixo um exemplo hipotético que demonstra como funcionaria a personalização. No exemplo, supõe-se um usuário cadastrado que demonstre um interesse maior por esportes, mais especificamente o futebol e o automobilismo. A figura 2 mostra como seria uma página padronizada para todos os usuários não-logados e aqueles recém-cadastrados, com pouca ou nenhuma informação coletada.



Figura 2 – Página inicial padronizada

A partir do momento em que o usuário realiza o login, qualquer modificação feita por ele permanecerá salva para os seus próximos acessos e cada passo dele no site será registrado pelo módulo de personalização.

Cada notícia terá associada a si no banco de dados não apenas a árvore de classificação hierárquica à qual esta pertença, indo do mais geral (por exemplo, “esportes”) até o mais específico (por exemplo, “Esporte Clube XYZ”). Cada vez que uma notícia é visualizada, as palavras-chave desta são armazenadas no módulo de personalização, para fins de ranqueá-las e, com isso, averiguar os conteúdos de preferências do usuário e, assim, retorná-los com maior prioridade sobre os demais.

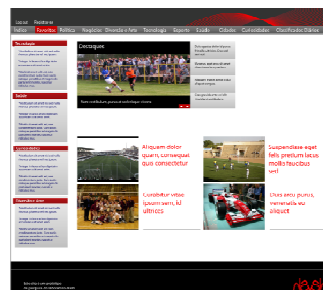


Figura 3 – Exemplo de página inicial personalizada

Com o decorrer dos acessos, a página sofrerá modificações de acordo com os tipos de conteúdos mais procurados. Se o usuário passa a maior parte do tempo acessando poucas notícias específicas de um assunto mais geral, as notícias sobre esse assunto mais geral

terão mais prioridade. Caso, ele demonstre principalmente por uma parte específica de um determinado assunto, acessando muitas notícias sobre este, novidades sobre este assunto terão um maior destaque em relação às demais. A figura 3 mostra como ficaria a página inicial do usuário tomado como exemplo após algum tempo, no qual já foi possível aferir suas preferências.

Desse modo, o usuário não precisa ter que ir até a seção “esportes” e, em seguida, “futebol” ou “automobilismo” para encontrar as notícias de seu interesse. Estas já seriam priorizadas pelo próprio perfil dele, economizando-o tempo. Caso o usuário demonstre um interesse ainda mais específico, como notícias relativa a um determinado clube de futebol, estas teriam um peso maior do que as demais notícias sobre futebol. Caso ele não nutra nenhum interesse específico por nenhum clube ou competição, as notícias relacionada ao tema “futebol” no geral receberiam o destaque por igual.

O objetivo é tornar as notícias de interesse do usuário disponíveis a ele de forma mais rápida e fácil, sem que ele tenha que procurá-las no site.

Quanto mais o site for acessado, mais informações serão disponibilizadas ao módulo de personalização e melhor será a triagem e o ranqueamento das categorias e, desta forma, mais personalizado o site se tornará, e, com isso, diferentes tipos de usuários podem ser atingidos pelo jornal.

4.3 Implementação

Uma vez definida a arquitetura do sistema, decidiu-se o formato em que esta seria construída: uma página de Web que funcionaria como um motor para a geração das páginas do site de notícia em questão. Ele teria duas interfaces: o site de notícias, para o usuário, e, para os jornalistas, um sistema de gerenciamento de conteúdo, que permitisse a adição, deleção e edição de notícias. Para que o motor independa do layout do site, foi especificado também o módulo de notícias.

Decidiu-se que o primeiro protótipo, para testar funcionamento do ranqueamento das categorias, seria feito utilizando o framework CakePHP [6]. Neste primeiro protótipo foi essencialmente priorizado a criação e os testes do módulo de personalização, não havendo, inclusive, preocupação em se fazer um layout. Pela simplicidade de uso, foi escolhido o MySQL como o banco de dados desta primeira versão.

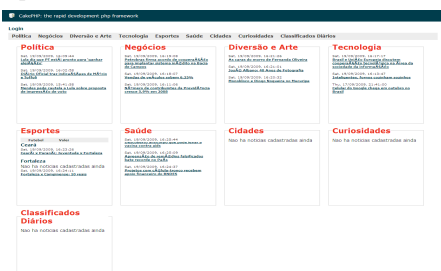


Figura 4 – Primeiro Protótipo

Foram criadas nove categorias de notícias (Política, Negócios, Diversão e Arte, Tecnologia, Esportes, Saúde, Cidades, Curiosidades e Classificados Diários), além de que a categoria “Esportes” era dividida em duas sub-categorias, “Vôlei” e “Futebol”, esta última, por sua vez, era dividida em duas sub-categorias, “Ceará” e “Fortaleza”.

Para funcionar como um proto-módulo de hierarquização foi decidido utilizar uma simples contagem de acessos às notícias

desses determinados temas. Quanto maior o número de acessos à determinada categoria, maior prioridade era dada a ela, podendo ela passar à frente das outras. Seguindo a árvore de hierarquia, um acesso em uma notícia classificada na sub-categoria “Ceará” também seria contada em sua categoria-mãe, “Futebol”, e na categoria-mãe desta, “Esportes”. Dessa maneira, as sub-categorias competiriam entre si pelo destaque na categoria-mãe, além de auxiliar esta a ganhar prioridades sobre as demais categorias.

Para a segunda versão do protótipo, decidiu-se abandonar o CakePHP em prol do Java Web, que facilitaria a implantação do conceito de ontologias, onde cada módulo seria uma classe em si, simplificando os relacionamentos entre elas. Além disso, cada módulo seria dividido em sub-classes, de acordo com a sua própria arquitetura interna, tornando a implementação mais simples. Por ora, a versão atual do protótipo tem apenas os módulos mais essenciais.

5. CONSIDERAÇÕES FINAIS

Com a queda nas vendas dos jornais impressos, um novo conceito de comercialização de notícias foi adotado pelos jornais em geral, que é o jornal on-line. A Internet traz diversas novas possibilidades para os jornais, contudo, possui alguns problemas que dificultam a sua adoção plena, tais como a forma de captação de lucro. Dentre outras soluções, o uso da publicidade é a mais aceita como candidata a fonte de renda para os jornais.

O objetivo deste estudo é buscar uma arquitetura capaz de tornar os sites de notícias mais atrativos para o grande público, adequando o jornalismo às novas formas de interação que a Internet trouxe. Como consequência da maior satisfação do público, maior será o lucro, tanto do meio publicitário quanto do próprio jornal, criando, assim, um ciclo virtuoso.

No presente momento do projeto, está se fazendo o levantamento dos requisitos necessários para o desenvolvimento de cada módulo do sistema e o aperfeiçoamento de um protótipo para testes.

6. AGRADECIMENTOS

Agradecemos ao CNPQ pelo apoio financeiro.

7. REFERÊNCIAS

- [1] S. Matsuura, “Circulação dos jornais cai 6,7% em abril.”, Disponível em: <<http://colunistas.ig.com.br/luisnassif/2009/06/01/a-queda-na-circulacao-dos-jornais/>> Acesso em: 17 jul 2009.
- [2] A. Andrews, “New York Times to decide how to charge for its website by August.”, Disponível em: <<http://www.telegraph.co.uk/finance/newsbysector/mediatechnologyandtelecoms/media/5786429/New-York-Times-to-decide-how-to-charge-for-its-website-by-August.html>> Acesso em: 17 jul 2009.
- [3] T. Doria, “O futuro dos jornais passa pelo Google e pela publicidade.”, Disponível em: <<http://www.tiagodoria.ig.com.br/2009/04/07/o-futuro-dos-jornais-passa-pelo-google-e-pela-publicidade/>> Acesso em: 17 jul 2009.
- [4] V. Schmitt e L. G. de Oliveira, “Personalização de Notícias - Uma edição de jornal para cada leitor”, Revista de Economía Política de las Tecnologías de la Información y Comunicación, vol. XI, nº. 1, 2009.
- [5] R. A. Falbo, F. B. Ruy, J. Pezzin e R. D. Moro, “Ontologias e Ambientes de Desenvolvimento de Software Semânticos”, Actas de las IV Jornadas Iberoamericanas de Ingeniería del Software e Ingeniería del Conocimiento, IIISIC'2004, Volumen I, pp. 277-292, Madrid, España, Noviembre 2004.
- [6] Cake Software Foundation Inc., CakePHP, Disponível em <<http://cakephp.org/>> Acesso em: 29 mar 2010.