

## Serviço de tradução de linguagens de marcação para a Internet

Luiz H. Z. Santana<sup>1</sup>, Diogo S. Martins<sup>1</sup>, Marcos Forte<sup>2</sup>, Wanderley Lopes de Souza<sup>1</sup>, Antônio Francisco do Prado<sup>1</sup>, Mauro Biajiz<sup>1</sup>, Luana Knoff<sup>1</sup>

<sup>1</sup>Departamento de Computação (DC) – Universidade Federal de São Carlos (UFSCar)  
Caixa Postal 676 – 13565-905 – São Carlos – SP

<sup>2</sup>Centro Universitário Fundação Santo André (FSA)  
Avenida Príncipe de Gales 821 – 09060-650 – Santo André – SP

{luiz\_santana, diogo\_martins, marcos\_forte, desouza, prado, mauro}@dc.ufscar.br, luana@comp.ufscar.br

**Abstract.** *Over the last few years, the Web's expansion has been accompanied by a growing number of mobile devices with Internet access capabilities. However, most Web-based content is still developed for presentation on personal computers, disregarding the limited resources of these mobile devices (e.g., processing, memory, power, screen, markup languages). This paper presents a markup language translation service provided by a content adaptation architecture for the Internet. The service was built based on Ontologies and Semantic Web Services, thus expanding the universe of Web pages accessible by mobile devices.*

**Resumo.** *Nos últimos anos a expansão da Web tem sido acompanhada por um número crescente de dispositivos móveis, cujas capacidades incluem o acesso à Internet. Entretanto, o desenvolvimento da maioria do conteúdo baseado na Web é ainda direcionado para apresentação em computadores pessoais e não leva em consideração os recursos limitados desses dispositivos móveis (e.g., processamento, memória, energia, tela, linguagens de marcação). Este artigo apresenta um serviço de tradução de linguagens de marcação provido por uma arquitetura de adaptação de conteúdo para a Internet. O serviço foi construído com base em Ontologias e Serviços Web Semânticos, ampliando dessa forma o universo de páginas Web acessíveis por dispositivos móveis.*

### 1. Introdução

Desde o seu advento, em meados da década de 90, a *Web* passou a fazer parte do cotidiano de diversos grupos e faixas etárias de usuários, exercendo uma grande influência na forma como informações são pesquisadas nos meios de comunicação e também nos processos de consumo de produtos e serviços [W3C]. Tal sucesso foi logo incorporado ao também notável crescimento do contingente de dispositivos móveis, tais como aparelhos celulares, *smart phones* e *Personal Digital Assistants (PDAs)*, que passaram em sua maioria a ser capazes de acessar a Internet. Apesar da participação crescente desses dispositivos como clientes da Internet, a autoria de conteúdos para a *World Wide Web (WWW* ou simplesmente *Web*) ainda é majoritariamente direcionada aos Computadores Pessoais (PCs), dificultando assim a entrega aos dispositivos móveis,

já que estes dispõem de recursos limitados (e.g., processamento, memória, energia, tela) quando comparados aos PCs.

A fim de lidar com esse problema pode-se adaptar os conteúdos, destinados aos PCs, para os dispositivos móveis e evitar assim a necessidade de desenvolvimento de conteúdos Web específicos para esses dispositivos. Geralmente tais processos de adaptação são realizados: (a) *no servidor de origem*, permitindo o fornecimento de conteúdos já adaptados aos dispositivos dos usuários, mas obrigando o servidor a conter várias páginas de um mesmo conteúdo; (b) *no dispositivo de usuário*, permitindo o desenvolvimento de processos específicos de adaptação em função dos recursos disponíveis no dispositivo, mas sobrecarregando-o com tais processos; (c) *num servidor dedicado*, o que alivia o servidor de origem e o dispositivo de usuário dos processos de adaptação, mas acrescentando um *overhead* principalmente em relação à comunicação. Dentre as adaptações, que buscam ampliar o universo de páginas Web acessíveis pelos dispositivos móveis, destaca-se a tradução de linguagens de marcação. Esse tipo de adaptação justifica-se pelo fato de grande parte dos dispositivos móveis possuírem *microbrowsers* capazes de apresentar páginas Web em linguagens distintas daquelas em que os conteúdos foram originalmente criados.

Este artigo apresenta um serviço de tradução de linguagens de marcação, o qual foi construído a partir do *Framework para Adaptação de Conteúdo na Internet (FACI)* [Claudino et al. 2005], que por sua vez foi desenvolvido para propiciar a diversos tipos de adaptação de conteúdo (e.g., compressão de dados, transcodificação de imagens, escaneamento de vírus, classificação e filtragem de conteúdo, inserção de publicidade, tradução de idiomas) serem realizados em servidores dedicados. Os serviços de adaptação são disponibilizados através de *Serviços Web Semânticos* [McIlraith et al. 2001, Hepp 2006], o que permite distribuí-los via uma rede de servidores, combinando-os automaticamente com base nas necessidades de cada adaptação. Neste artigo o termo *Serviços Web Semânticos* refere-se ao uso de tecnologias da *Web Semântica* para a descrição, a descoberta e a composição de *Serviços Web*. Tais requisitos são orientados por uma política de adaptação que considera a semântica dos serviços, apoiando-se em perfis de contexto de entrega e perfis de serviço, ambos baseados em *Ontologias* [Guarino 1998].

A seqüência deste artigo está organizada da seguinte forma: a seção 2 introduz o problema de tradução de linguagens de marcação, discutindo possíveis soluções para o mesmo; a seção 3 trata da política de adaptação empregada, focando no contexto de entrega e nas regras de adaptação; a seção 4 discorre sobre o serviço de tradução de linguagens de marcação, expondo as vantagens do emprego de *Serviços Web Semânticos*, mostrando como esse serviço insere-se no contexto de uma arquitetura de adaptação de conteúdo para a Internet e como o mesmo foi construído a partir do FACI; a seção 5 descreve os testes realizados e os resultados obtidos durante a avaliação desse serviço; a seção 6 versa sobre trabalhos relacionados ao tema alvo deste artigo, comparando-os ao trabalho aqui apresentado; finalmente a seção 7 tece algumas conclusões relativas à pesquisa desenvolvida e aponta para trabalhos futuros.

## 2. Tradução de linguagens de marcação

Atualmente coexistem diversas linguagens de marcação para a Web, destacando-se: *HyperText Markup Language (HTML)*, *Extensible HTML (XHTML)*, *XHTML Mobile*

*Profile (XHTML-MP)*, *Wireless Markup Language (WML)* e *compact HTML (cHTML)*. Porém, os *microbrowsers* dos dispositivos móveis de acesso restringem-se a suportar um número pequeno de elementos desse conjunto e, muitas vezes, apenas um deles.

Uma solução para esse problema, que visa ampliar o universo de páginas Web acessíveis pelos dispositivos móveis, é a conversão do conteúdo original, descrito numa determinada linguagem de marcação, para uma linguagem que seja suportada pelo dispositivo do usuário. Neste sentido o Fórum *Wireless Application Protocol (WAP)*<sup>1</sup>, em conjunto com outros parceiros, propôs as seguintes abordagens de conversão [Metter e Colomb 2000]: (a) *autoria em XML*, na qual o conteúdo é marcado em XML e são aplicadas transformações no documento, a fim de obter versões nas respectivas linguagens de marcação suportadas pelos dispositivos; (b) *ferramentas de autoria*, são capazes de gerar código para diversas linguagens de marcação demandando, posteriormente, a publicação de várias versões do conteúdo; (c) *conversão em lote*, na qual um conjunto de páginas Web pré-existentes é convertido em lote (e.g., de HTML para WML); (d) *conversão sob demanda*, em que as páginas são convertidas, no momento do acesso, de acordo com as características do dispositivo.

A *autoria em XML* e as *ferramentas de autoria* são abordagens que influenciam no processo de autoria e, embora permitam ao autor o máximo de controle sobre o conteúdo entregue, não abrangem o tratamento de conteúdos já publicados.

A *conversão em lote* demanda, para cada linguagem na qual será disponibilizado o conteúdo, o re-processamento de uma grande massa de documentos. Assim como nas *ferramentas de autoria*, essa abordagem exige a presença, no servidor de origem, de várias versões de uma mesma página e de um mecanismo para selecionar, com base nas características do dispositivo, a versão na linguagem de marcação apropriada.

A *conversão sob demanda* dispõe das seguintes vantagens sobre as demais abordagens: não interfere no processo de autoria; permite usar um servidor dedicado, aliviando o servidor de origem; permite usar *cache*, evitando a repetitiva conversão de uma mesma página. Entretanto, essa abordagem exige que as informações de contexto de entrega sejam referenciadas com frequência, para que as conversões sejam guiadas.

Em relação à apresentação, as características de uma página Web são tradicionalmente classificadas em quatro grupos [Glover e Davies 2005]: *conteúdo*, formado pelas informações que são apresentadas ao usuário; *estilo*, composto pelos atributos de formatação (e.g., cores, fontes); *layout*, constituído pelos relacionamentos geométricos e topológicos entre os diversos componentes da página; *estrutura*, caracterizada pelos relacionamentos entre páginas ou entre âncoras de uma página.

Dependendo da linguagem de marcação, na qual foi criada uma determinada página Web, esses grupos de características estarão presentes em diferentes graus. Por exemplo, documentos HTML inibem a separação entre conteúdo e estilo, enquanto documentos XHTML incentivam essa distinção. Conseqüentemente, a mera conversão da linguagem de marcação, desconsiderando as variações nas características de apresentação, pode levar a resultados insatisfatórios. Para lidar com esse problema, algumas abordagens têm sido propostas visando a modificação da apresentação:

---

<sup>1</sup> <http://www.wapforum.org/>

- (a) *Sumarização*, que disponibiliza ao usuário resumos dos tópicos da página original. Um jornal *online*, por exemplo, pode apresentar apenas os títulos das notícias presentes na página principal e oferecer *links* para as notícias completas. Dessa forma, a página é afetada tanto em seu conteúdo como em sua estrutura [Buyukkokten et al. 2002];
- (b) *Thumbnails*, que apresenta uma fotografia da página como um todo e permite que o usuário escolha regiões a serem ampliadas. Para tal faz uso de mapas de imagens, ou seja, divide a fotografia em diversas regiões e, para cada região, oferece um link para uma versão ampliada, afetando assim a estrutura da página [Lam e Baudisch 2005];
- (c) *Mediação por prioridades*, em que são identificadas as seções de uma mesma página ou conjuntos de páginas, que sejam mais relevantes ao usuário, de acordo com alguma métrica. Por exemplo, pode-se entregar aos dispositivos apenas as seções que contêm os *links* mais acessados pelos usuários [Song et al. 2004, Bar-Yossef e Rajagopalan 2002].

A técnica de sumarização é a mais fácil de ser implementada em relação às outras opções, porém cria um nível adicional na estrutura de navegação. Para *Websites* com elevado número de páginas, pode gerar índices que não cabem na tela do dispositivo móvel e exigem muita rolagem. A abordagem de *thumbnails*, por sua vez, é a mais amigável ao usuário, já que apresenta a página em sua forma original. Entretanto, é dependente do suporte a recursos de mapas de imagens nos *browsers* dos dispositivos.

Finalmente, a mediação por prioridades pode propiciar maior satisfação ao usuário, pois é uma forma de adaptação às suas preferências. Porém, exige que as informações sobre o contexto de entrega possuam uma boa amostragem. As prioridades tratadas por essa técnica podem ser coletadas em nível local ou global.

A coleta de prioridades globais busca minerar padrões de similaridade entre blocos lógicos (parágrafos, tabelas, camadas, etc.) em todas as páginas de um mesmo Website [Bar-Yossef e Rajagopalan 2002]. Os blocos similares de maior frequência recebem baixa prioridade, pois comumente são gerados por *templates* e representam um conteúdo de menor importância (cabecinhos, informações de *copyright*, barras de navegação). Já os blocos com menor frequência tendem a representar as informações mais relevantes das páginas, sendo atribuído aos mesmos uma maior prioridade.

As prioridades locais são atribuídas a blocos lógicos de uma mesma página, tais como parágrafos, tabelas e camadas [Song et al. 2004]. A partir da representação em árvore do documento associa-se um valor de prioridade a cada nó, baseando-se em alguma métrica (e.g., os hábitos de navegação dos usuários).

De forma a atender eficientemente esses dois requisitos, isto é, tradução da linguagem de marcação e modificação da apresentação, este trabalho propõe um serviço de tradução sob demanda de páginas Web, que modifica a apresentação via uma técnica de mediação por prioridades locais, que por sua vez é orientada metricamente pelo número de acessos do usuário aos blocos lógicos.

### 3. Política de adaptação

A política de adaptação define quais serviços de adaptação serão oferecidos, quais adaptadores locais ou remotos executarão as adaptações e quando estes serão solicitados. Para tal essa política utiliza, conjuntamente, as informações relativas ao contexto de entrega e as regras de adaptação.

O contexto de entrega consiste de um conjunto de atributos que caracteriza as capacidades dos dispositivos de acesso, os dados pessoais e as preferências dos usuários, as condições da rede de acesso, as características dos conteúdos requisitados e as resoluções contratuais entre o provedor de serviços e o usuário final [Souza et al. 2002]. Essas classes de informações são definidas, respectivamente, nos seguintes perfis: dispositivo, usuário, rede, conteúdo e *Service Level Agreement (SLA)*.

Dentre os padrões adotados para a descrição de perfis destacam-se: *User Agent Profile (UAProf)*<sup>2</sup>, *Composite Capabilities/Preferences Profile (CC/PP)*<sup>3</sup> e *Wireless Universal Resource File (WURFL)*<sup>4</sup>, sendo esse último um extenso banco de dados, em XML, contendo características de dispositivos. Entretanto, nenhum desses três padrões descreve todas as características do contexto de entrega.

Para descrever características de outras classes de informações do contexto de entrega, tais como os perfis de usuário e de conteúdo, há uma carência de vocabulários específicos. Para atender a essa necessidade, é comum o emprego de *Resource Description Framework (RDF)* [Forte et al. 2006] na descrição de perfis. Porém essa alternativa limita as possibilidades de inter-relacionamentos entre os atributos de contexto, uma vez que permite representá-los somente via taxonomias.

Para tornar os perfis mais expressivos, pode-se usar linguagens para a definição de ontologias, tais como a *Web Ontology Language (OWL)*<sup>5</sup>. Assim como RDF, OWL é considerada uma tecnologia-chave para a Web Semântica, por ser uma linguagem para a representação de conhecimento, que permite a especificação explícita e compartilhada de conceitos e relações de um determinado domínio de aplicação [Guarino 1998]. Dentre as razões, que tornam o uso de ontologias adequado aos sistemas que lidam com informações contextuais, destacam-se [Wang et al. 2004]: permitem compartilhar o conhecimento entre diversos sistemas; permitem a agentes inteligentes argumentarem sobre informações de contexto; permitem a interoperabilidade entre sistemas que não foram projetados para esse fim.

Em função dessas vantagens, neste trabalho optou-se pela descrição dos perfis através de ontologias. Na Figura 1 é ilustrado o modelo de ontologia para o perfil de conteúdo, sendo que *Browser\_Accept*, *Others*, *Location*, e *Content\_Info* são sub-classes de *Content*. Maiores informações relativas a esse modelo, assim como pertinentes aos modelos de ontologias para os demais perfis, estão disponíveis em [Forte et al. 2006].

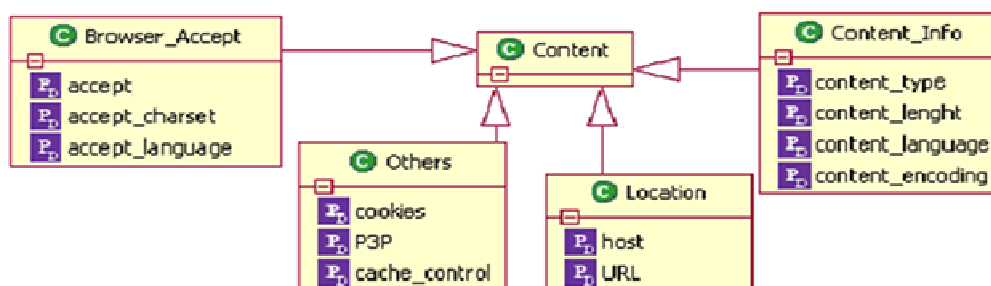


Figura 1. Modelo de ontologia para o perfil de conteúdo

<sup>2</sup> [http://w3development.de/rdf/uaprof\\_repository/](http://w3development.de/rdf/uaprof_repository/); <sup>3</sup> <http://www.w3.org/Mobile/CCPP/>; <sup>4</sup> <http://wurfl.sourceforge.net/>

<sup>5</sup> <http://www.w3.org/2004/OWL/>; <sup>6</sup> W3C Semantic Web Activity [<http://www.w3.org/2001/sw/>]

Combinando as informações de contexto de entrega, descritas nos perfis de dispositivo, usuário, rede, conteúdo e SLA, com as regras de adaptação, o mecanismo de decisão do FACI [Claudino 2005] determina os serviços necessários para a execução de uma determinada adaptação.

A fim de tirar proveito da formalização dos perfis via ontologias e visando facilitar a composição de serviços de adaptação, as regras de adaptação foram especificadas em *Semantic Web Rule Language (SWRL)*<sup>7</sup>. Essa linguagem é baseada em cláusulas de *Horn* e está fortemente associada aos Serviços Web Semânticos os quais, neste trabalho, são empregados na implementação dos serviços de adaptação.

#### 4. Serviço de tradução de linguagens de marcação

Serviços Web são elementos computacionais auto-descritivos e independentes de plataforma, que suportam a composição de aplicações distribuídas de uma forma rápida e com baixo custo [Papazoglou 2003]. Em relação à adaptação de conteúdo, o uso dessa tecnologia possibilita tanto a distribuição dos serviços de adaptação, através de uma rede de servidores dedicados conectados à Internet, quanto a composição desses serviços para a realização de adaptações mais complexas.

Durante a tradução de linguagens de marcação, muitas vezes torna-se necessário empregar mais de um serviço para realizar essa tarefa. Por exemplo, para atender às restrições de um determinado dispositivo, além dos serviços de tradução e modificação de apresentação, pode ser necessário também um serviço de adaptação multimídia para redimensionar imagens do conteúdo. Nesse caso, uma técnica de composição de Serviços Web Semânticos pode fazer uso da política de adaptação para escolher e combinar os serviços necessários.

Serviços em geral devem ser descritos, descobertos e acessados, sendo que os Serviços Web tradicionais permitem a realização dessas três tarefas através do emprego de tecnologias baseadas em XML. Adicionalmente, os Serviços Web Semânticos tiram proveito das tecnologias de representação de conhecimento, relacionadas à Web Semântica, para agregar um maior poder de expressão a essas tarefas. Isso permite o uso de algoritmos apropriados para automatizar a descoberta e a composição de serviços. Dentre as várias abordagens [Davies et al. 2006] para a implementação de Serviços Web Semânticos, neste trabalho foi adotado a *Ontology Web Language-Service (OWL-S)*<sup>8</sup>, que é uma proposta de padrão submetida ao *World Wide Web Consortium (W3C)*.

OWL-S é uma ontologia para Serviços Web Semânticos, que possibilita uma rápida migração das descrições sintáticas da *Web Services Description Language (WSDL)* para descrições semânticas da OWL. OWL-S é organizada em três sub-ontologias: *ServiceProfile*, que descreve as características e capacidades do serviço, ou seja, as transformações que esse serviço é capaz de realizar; *ProcessModel*, que especifica o protocolo de interação com o serviço; e *ServiceGrounding*, que provê meios para que o contexto de entrega, expresso através de ontologias, possa ser enviado via mensagens *Simple Object Access Protocol (SOAP)*<sup>9</sup>.

---

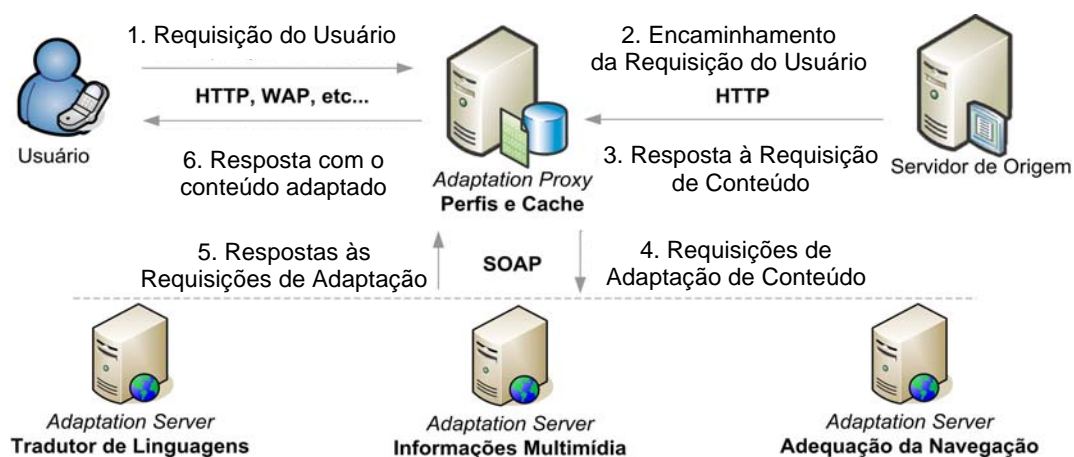
<sup>7</sup> <http://www.w3.org/Submission/SWRL/>; <sup>8</sup> <http://www.w3.org/Submission/OWL-S/>

<sup>9</sup> <http://www.w3.org/Submission/OWL-S/>

A OWL-S é usada neste artigo para a especificação do perfil de serviço, que descreve as características de um serviço de adaptação, incluindo as condições necessárias para a sua execução. Essas características podem ser divididas em duas classes: as diretamente associadas ao processo de adaptação, representadas pelas *entradas* e *saídas*; e as de apoio, representadas pelas *pré-condições* e *efeitos*, que auxiliam na decisão quanto à execução (ou não) de um determinado serviço. O uso dessa abordagem no projeto do serviço de tradução de linguagens de marcação permitiu, além da distribuição desse serviço pela Internet, a automatização da descoberta e da composição de serviços de acordo com a política de adaptação.

#### 4.1 Arquitetura de adaptação de conteúdo

Conforme ilustrada na Figura 2, a arquitetura de adaptação de conteúdo para a Internet, originalmente concebida para suportar o protocolo *Internet Content Adaptation Protocol* (ICAP) [Souza et al. 2002], foi estendida para também suportar o protocolo SOAP. Isso possibilita o uso de Serviços Web Semânticos e a implementação distribuída do serviço de tradução de linguagens de marcação via um conjunto de servidores dedicados. O servidor de tradução foi implementado reutilizando-se o FACI [Claudino 2005], ao qual foram incorporadas novas funcionalidades relativas aos Serviços Web Semânticos.



**Figura 2. Arquitetura de adaptação de conteúdo para a Internet**

O fluxo de mensagens, representado na Figura 2, tem início quando o Usuário requisita um conteúdo ao Servidor de Origem (1). Essa requisição é interceptada pelo *Adaptation Proxy*, que realiza uma análise preliminar da mesma considerando a política de adaptação e o conteúdo do cache. Se o conteúdo requisitado não estiver no cache, o *Adaptation Proxy* encaminha a requisição ao Servidor de Origem (2), o qual responde com o conteúdo requisitado (3). De posse desse conteúdo, o *Adaptation Proxy* realiza uma análise definitiva para determinar a necessidade (ou não) do serviço de tradução. Em caso positivo, este decide quais *Adaptation Servers* devem ser requisitados, para o fornecimento do serviço de tradução, e estabelece uma ordem para o envio das requisições de adaptação aos *Adaptation Servers* (4). Estes respondem às requisições do *Adaptation Proxy*, enviando-lhe os conteúdos adaptados (5). Caso essas respostas indiquem que as adaptações foram bem-sucedidas, a combinação das mesmas é armazenada no cache e encaminhada ao usuário (6). Toda a comunicação entre o

Adaptation Proxy e os Adaptation Servers é realizada através do protocolo SOAP, uma vez que cada Adaptation Server corresponde a um Serviço Web Semântico.

#### 4.2. Framework de Adaptação de Conteúdo

O FACI é composto pelos pacotes *Adaptation Proxy* e *Adaptation Server* [Claudino 2005]. Uma vez que esse framework não foi originalmente concebido para suportar Serviços Web Semânticos, novos componentes foram adicionados ao pacote *Adaptation Proxy* [Forte et al. 2006] para viabilizar o uso de tais serviços. A Figura 3 apresenta o modelo de componentes do Adaptation Proxy estendido. O componente *Content Transfer Protocol* é responsável pela comunicação com o servidor de origem e o dispositivo do usuário, podendo lidar com diferentes protocolos (e.g., HTTP, WAP). O componente *ProxyManager* é responsável pela gerência de cada requisição e pela consulta ao componente *Cache* em busca de adaptações previamente realizadas. A comunicação com o pacote *Adaptation Server* é de responsabilidade do componente *Callout Protocol Client* que, nesse caso, utiliza o protocolo SOAP.

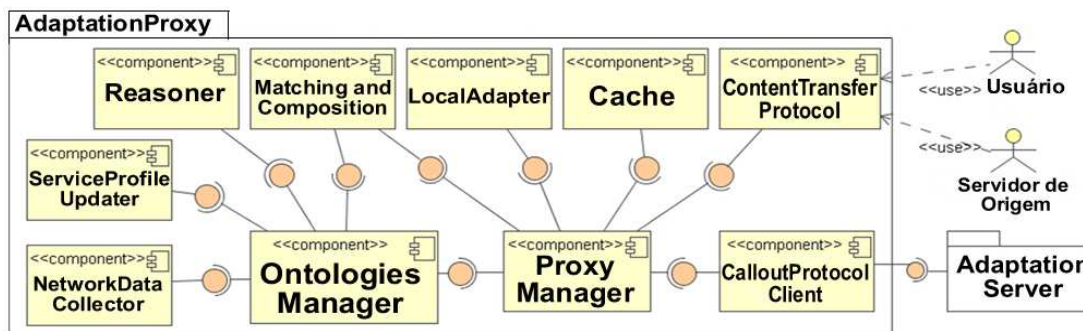


Figura 3. Modelo de componentes do *Adaptation Proxy* estendido

O tratamento da política de adaptação é realizado pelo componente *Ontologies Manager*. Nesse sentido, o componente *Network Data Collector* atua como agente de atualização do perfil de rede e o componente *ServiceProfileUpdater* trata do perfil de SLA.

As informações da política de adaptação são recuperadas pelo componente *Matching and Composition* através de consultas em *SPARQL*<sup>10</sup>. Tais informações são usadas pelo componente *Reasoner* para decidir a necessidade ou não de uma determinada adaptação. Em caso positivo, as pré-condições e os efeitos dos serviços de adaptação [Forte et al. 2006], descritos nos perfis de serviço, são verificados para indicar se é preciso ou não mais de um serviço para realizar essa adaptação. Em caso afirmativo, esses serviços são determinados e uma ordem é estabelecida para as invocações dos mesmos.

A composição dos serviços pode ser estática ou dinâmica. No primeiro caso deve-se especificar previamente, por meio de um fluxo de trabalho, como os serviços interagem entre si para oferecer o conjunto de funcionalidades de uma determinada adaptação. Já na composição dinâmica os serviços são selecionados automaticamente de acordo com a política de adaptação.

<sup>10</sup> <http://www.w3.org/TR/rdf-sparql-query/>



Para determinar os serviços necessários são utilizadas regras de adaptação, descritas em SWRL, nas ontologias do perfil de serviço. Tal abordagem permite que novas regras sejam inseridas, pelo mecanismo de inferência, na base de conhecimento, minimizando assim o impacto de modificações sobre os componentes responsáveis pela política de adaptação.

Para viabilizar tanto o tratamento dos aspectos de apresentação da página Web quanto o uso de um segundo nível de cache, novos componentes foram adicionados ao pacote *Adaptation Server* do FACI. A Figura 4 apresenta o modelo de componentes do *Adaptation Server* estendido. O componente *Callout Protocol Server* é responsável pela comunicação com o pacote *Adaptation Proxy*. A gerência das conexões é realizada pelo componente *ServerManager*. O componente *Cache* armazena os conteúdos adaptados em uma representação intermediária e, posteriormente, a adaptação adequada a cada contexto de entrega é realizada pelo componente *Content Adapter*. Finalmente, o componente *Presentation Adapter* trata dos aspectos de apresentação do conteúdo que sofreu adaptação.

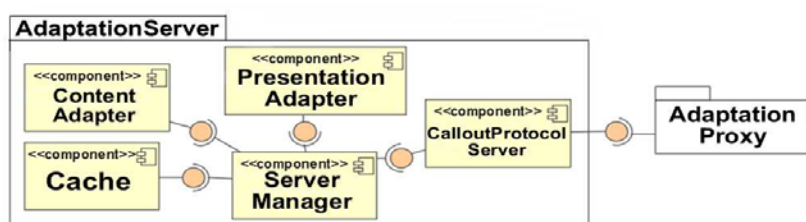


Figura 4. Modelo de componentes do *Adaptation Server* estendido

Cada serviço de adaptação fica sob a responsabilidade de um *Adaptation Server*. Para o serviço de tradução de linguagens de marcação foram definidos três *Adaptation Servers*: *Adaptador de Informações Multimídia*, para o tratamento de imagens; *Tradutor de Linguagens de Marcação*, para a conversão de linguagens de marcação; *Adaptador de Navegação*, para o tratamento da apresentação por mediação de prioridades.

O *Adaptador de Informações Multimídia* usa o componente *ContentAdapter* para realizar transformações de escala, de cores ou o eventual descarte de imagens. Já o componente *PresentationAdapter* é usado para a transcodificação entre formatos (e.g., JPEG, PNG, GIF, WBMP) por meio da API *Java Advanced Imaging (JAI)*. Após ser adaptada, cada imagem é enviada ao *Adaptation Proxy* para que seja encaminhada ao dispositivo do usuário.

O *Tradutor de Linguagens de Marcação* envolve o processo ilustrado na Figura 5a. Nesse serviço adotou-se XHTML, como representação intermediária das páginas a serem traduzidas. Se as páginas do servidor de origem já estiverem nesse formato, estas serão processadas diretamente pelo componente *ContentAdapter*, a fim de que este possa extrair as informações a serem armazenadas no perfil de conteúdo. Se as páginas estiverem em HTML, um passo adicional é necessário para a conversão das mesmas para XHTML. Em ambos os casos as páginas XHTML são armazenadas no componente *Cache*. Finalmente, o componente *PresentationAdapter* realizará a tradução das páginas para as linguagens determinadas pela política de adaptação.

Os aspectos de apresentação são tratados pelo *Adaptador de Navegação*, cujo processo está representado na Figura 5b. Esse processo guia-se pela mediação das

prioridades atribuídas aos blocos lógicos da página (e.g., tabelas, parágrafos, camadas). Para tal, as páginas que são mantidas no componente *Cache* recebem marcações especiais, a fim de que os blocos possam ser identificados. No perfil de conteúdo é associada uma prioridade ao identificador de cada bloco lógico. A métrica adotada, para a definição de prioridades, é a importância do bloco, a qual é calculada como  $(qt\_acesso + fator\_importancia)/qt\_links$ , onde: *qt\_acesso* é o número total de acessos que um bloco lógico obteve; *fator\_importancia* é determinado em função da quantidade de informação (em caracteres) do bloco, a sua posição na página e se este contém imagens; e *qt\_links* representa a quantidade de links encerrados num bloco. A prioridade de um bloco aumenta à medida que seus *links* são mais acessados ao longo do tempo.

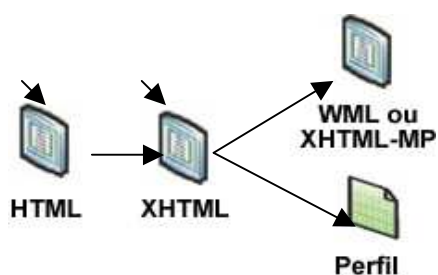


Figura 5a. Processo de tradução de páginas HTML.



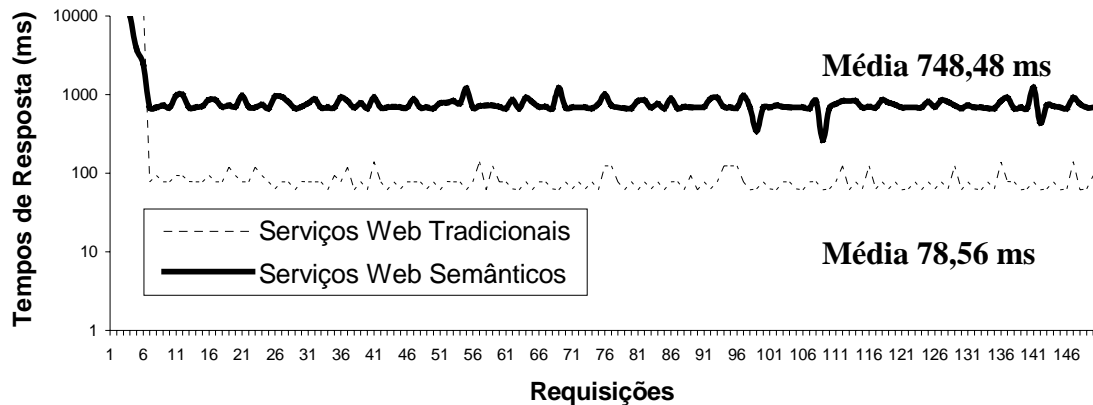
Figura 5b. Seqüência de adaptação da navegação.

## 5. Avaliação do serviço de tradução

Para avaliar o desempenho, do serviço de tradução de linguagens de marcação, foi implementada uma configuração de rede envolvendo 04 computadores Pentium 4-2.5 Ghz Dual Core com 2 GB de memória RAM, todos usando Windows XP. Num computador foi instalado o Adaptation Proxy, com uma base de dados contendo os acessos ao proxy do laboratório de graduação do Departamento de Computação (DC) da Universidade Federal de São Carlos (UFSCar), enquanto nos demais foram instalados os Adaptation Servers. Os testes consistiram no cálculo do tempo de resposta para cada requisição, em diferentes configurações do Adaptation Proxy, enfatizando a análise de desempenho do serviço de tradução quando submetido a um grande número de acessos.

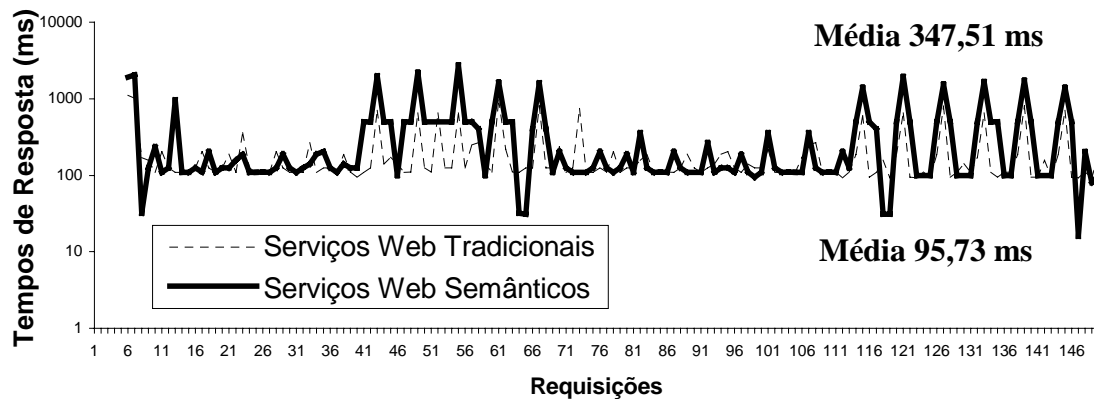
Durante 09 horas foram coletadas milhares de requisições ao proxy. Somente as 150 primeiras estão ilustradas nos gráficos das Figuras 6 e 7, já que o comportamento presente nesse conjunto de requisições repete-se ao longo do tempo. Conseqüentemente, em cada gráfico, os padrões das curvas Serviços Web Semânticos e Serviços Web Tradicionais são os mesmos. Em cada uma dessas figuras estão representados os testes realizados com o uso de Serviços Web tradicionais e com o uso de Serviços Web Semânticos, este último via *API OWL-S Mindswap*<sup>11</sup>. Os testes visualizados na Figura 6 são com o cache do serviço desativado enquanto que na Figura 7 são com o cache do serviço ativado.

<sup>11</sup> OWL-S API – Mindswap Maryland University – <http://www.mindswap.org/>



**Figura 6. Tempos de resposta às requisições com o cache desativado**

De acordo com a Figura 6, o uso de Serviços Web Semânticos torna o tempo de resposta às requisições em média 10 vezes maior que o tempo de resposta com o uso de Serviços Web tradicionais. Esse resultado já era esperado porque o emprego de Serviços Web Semânticos aumenta a complexidade do processo de adaptação de conteúdo como um todo e porque a API usada possui um baixo desempenho, já que foi desenvolvida para fins acadêmicos. Entretanto, essa perda quantitativa no tempo de resposta é largamente compensada pelo ganho qualitativo propiciado por essa complexidade adicional, ou seja: a capacidade de realizar inferências sobre o contexto de entrega para determinar os serviços necessários à execução de uma determinada adaptação; a busca automática desses serviços; a composição também automática desses serviços [Yao 2006].



**Figura 7. Tempos de resposta às requisições com o cache ativado**

De acordo com a Figura 7, o uso do cache acarretou num pequeno aumento no tempo de resposta, quando Serviços Web tradicionais são empregados, e numa redução considerável desse tempo, quando os Serviços Web Semânticos são utilizados. Nesse sentido, foi usada a mesma escala logarítmica nos gráficos 6 e 7, o que aproxima as duas curvas em 7, para ressaltar que o simples emprego de cache diminui significativamente a perda de desempenho causada pelo uso de Ontologias e Serviços Web Semânticos. O aumento é devido ao *overhead*, introduzido pelo acesso ao disco durante a inserção e a recuperação de dados no cache, e a redução é devida à presença de conteúdos já adaptados no cache, evitando assim a repetição de todo o processo de adaptação. Outro aspecto a ser ressaltado é a presença de dois conjuntos de picos de tempos de resposta,

que estão nas faixas 43 a 64 e 115 a 140. Esses picos são decorrentes do acesso a um conjunto de páginas ainda não presentes no cache. Já que é comum o uso de cache em *proxies*, esses resultados justificam ainda mais o emprego de Serviços Web Semânticos para a adaptação de conteúdo na Internet.

## 6. Trabalhos correlatos

Embora a necessidade de adaptação de conteúdo para dispositivos móveis não seja um problema recente, a ausência de soluções que dominem o mercado tem propiciado o surgimento de um grande número de trabalhos nessa área. Em geral esses trabalhos buscam o desenvolvimento de *proxies* dedicados à adaptação de conteúdo, ou buscam apenas soluções para a apresentação desse conteúdo nos dispositivos móveis.

Na primeira categoria destaca-se o trabalho [Laakko e Hiltunen 2005], onde o proxy leva em consideração as características de dispositivo, de sistema e de usuário para a realização da adaptação de conteúdo. Uma proposta hierárquica de adaptação de navegação é apresentada, mas que não considera a importância dos diferentes blocos lógicos que compõem a página Web a ser adaptada. Já no trabalho [Lemlouma e Layaida 2004] o proxy usa Serviços Web tradicionais para a adaptação de conteúdo. Apesar da proposta de adaptação de navegação ser baseada na técnica de sumarização, esta também não considera a importância dos blocos lógicos. Comparando esses trabalhos ao serviço de tradução de linguagens de marcação proposto, este apresenta as seguintes vantagens: o uso conjunto de perfis, regras e tecnologias da Web Semântica, aumentando a robustez e facilitando a manutenção do sistema; uma melhor qualidade de adaptação de navegação, na medida em que são consideradas as preferências do usuário para a definição da importância dos blocos lógicos de uma página Web.

Na segunda categoria destaca-se o trabalho [Pashtan et al. 2003], onde o conteúdo é adaptado em duas etapas: na primeira é gerada uma estrutura intermediária em árvore do conteúdo, enquanto que na segunda etapa esse conteúdo é traduzido para a linguagem de marcação determinada pela política de adaptação. No serviço de tradução proposto também é gerada uma estrutura intermediária, porém no padrão XHTML, facilitando assim a segunda etapa, que no caso é a tradução para WML ou XHTML-MP.

No trabalho [Gu et al. 2004] são apresentadas duas abordagens para a entrega de conteúdo adaptado: na primeira, denominada adaptação parcial, a importância dos blocos lógicos de uma página Web é definida no processo de autoria, rotulando-se esses blocos; na segunda, denominada negociação de versão, a importância dos blocos lógicos é definida dinamicamente pelo sistema, através da adição de marcações de importância nesses blocos, eliminando assim a necessidade de um perfil de conteúdo, o que melhora o desempenho do sistema como um todo. Essa segunda abordagem é semelhante a deste artigo, contudo no serviço de tradução proposto a marcação é usada como uma ligação com o perfil de conteúdo, onde diferentes informações semânticas (não apenas a importância) podem ser definidas. Este trabalho, assim como os trabalhos [Laakko e Hiltunen 2005, Lemlouma e Layaida 2004, Gu et al. 2004], não faz qualquer análise quantitativa dos serviços.

O trabalho [Baluja 2006] usa a abordagem de thumbnails, aliada a conceitos de aprendizagem de máquinas, na adaptação de navegação. Tais conceitos possibilitam que, na entrega do conteúdo adaptado, as preferências do usuário possam ser melhor

atendidas, quando comparadas ao serviço de tradução proposto, outorgando assim um maior grau de satisfação ao mesmo. Este trabalho avalia o tempo de entrega de uma página, enquanto o trabalho proposto foca na análise do tempo de adaptação da página.

## 7. Conclusão

O serviço de tradução de linguagens de marcação, apresentado neste artigo, é flexível, *escalável* e de fácil manutenção, na medida em que foi construído com base num framework para adaptação de conteúdo na Internet. O FACI foi desenvolvido para propiciar que a carga de um serviço de adaptação possa ser distribuída por uma rede de servidores dedicados e para permitir que novos serviços de adaptação possam ser facilmente construídos a partir do mesmo. Esse framework foi estendido para suportar a combinação automática de serviços de adaptação, via o emprego de Ontologias e Serviços Web Semânticos, de acordo com o contexto de entrega. A descrição detalhada do FACI estendido e a sua avaliação qualitativa foram objetos do artigo referenciado [Forte 2003]. O objetivo desse novo artigo foi mostrar como construir um serviço de adaptação de conteúdo a partir do FACI estendido e avaliar o impacto do uso de Ontologias e Serviços Web Semânticos no desempenho do serviço construído.

Pode-se apontar algumas melhorias a serem introduzidas no serviço de tradução proposto: usar APIs mais eficientes para a manipulação das informações do contexto de entrega; considerar outras informações do contexto de entrega para a adaptação de navegação; empregar critérios mais sofisticados para o cálculo das importâncias na técnica de mediação por prioridades; e incluir serviços de adaptação que possibilitem também o tratamento de áudio e vídeo presentes na Web. Todas essas melhorias visam aumentar o grau de satisfação do usuário em relação ao conteúdo adaptado.

## Referências

- Baluja, S. (2006) "Browsing on small screens: recasting web-page segmentation into an efficient machine learning framework". Anais da 15<sup>o</sup> International Conference on World Wide Web, pp. 33–42.
- Bar-Yossef, Z. e Rajagopalan, S. (2002) "Template detection via data mining and its applications". Anais da 11<sup>o</sup> International Conference on World Wide Web, pp. 580–591.
- Buyukkokten, O. et al. (2002) "Efficient Web Browsing on Handheld Devices Using Page and Form Summarization". Transactions on Information Systems, vol. 20, no.1, pp. 82–115.
- Claudino M., Souza, W.L. e Prado, A.F. (2005) "Um framework baseado em componentes para o domínio de adaptação de conteúdo na Internet". Anais do XIX Simpósio Brasileiro de Engenharia de Software, pp. 88-103.
- Davies, J., Studer, R. e Warren, J.P. (2006) "Semantic Web Technologies - Trends and Research in Ontology-based Systems". John Wiley & Sons.
- Forte, M., Souza, W.L., e Prado, A.F. (2006) "Utilizando ontologias e serviços web na computação ubíqua". Anais do XX Simpósio Brasileiro de Engenharia de Software, pp. 287-302.

- Glover, T. e Davies, J. (2005) “Integrating device independence and user profiles on the web”, *BT Technology Journal*, vol. 23, no. 3, pp. 239–248.
- Guarino, N. (1998) “Formal ontology in information systems”, *Anais da International Conference on Formal Ontologies in Information Systems*, pp. 3-15.
- Gu, W. e Helal, A.S. (2004) “An XML Based Solution to Delivering Adaptive Web Content for Mobile Clients”. *Anais do Symposium on Performance Evaluation of Computer and Telecommunication Systems*.
- Hepp, M. (2006) “Semantic Web and Semantic Web Services: Father and Son or Indivisible Twins?”, *IEEE Internet Computing*, vol. 10, no 2, pp. 85–88.
- Laakko, T. e Hiltunen, T. (2005) “Adapting Web Content to Mobile User Agents”. *IEEE Internet Computing*, vol. 9, no. 2, pp. 46–53.
- Lam, H. e Baudisch, P. (2005) “Summary thumbnails: readable overviews for small screen web browsers”. *Anais da Conference on Human Factors in Computing Systems*, pp. 681–690.
- Lemlouma, T. e Layaida, N. (2004) “Context-aware adaptation for mobile devices”. *Anais da 2004 IEEE International Conference on Mobile Data Management*, pp. 106–111.
- McIlraith, S.A., Son, T.C. e Zeng, H. (2001) “Semantic Web services”, *IEEE Intelligent Systems*, vol. 16, no. 2, pp. 46–53.
- Metter, M. e Colomb, R. (2000) “WAP enabling existing HTML applications”, *Anais da First Australasian User Interface Conference*, pp. 49–57.
- Papazoglou, M.P. (2003) “Service-oriented computing: concepts, characteristics and directions” *Anais da Conference on Web Information Systems Engineering*, pp. 3–12.
- Pashtan, A., Kollipara, S. e Pearce, M. (2003) “Adapting content for wireless Web services”. *IEEE Internet Computing*, vol. 7, no. 5, pp. 79–85.
- Song, R. et al. (2004) “Learning block importance models for web pages”. *Anais da 13° International Conference on World Wide Web*, pp. 203–211.
- Souza, W.L. et al. (2002) “Adaptação de conteúdo de mensagens HTTP baseada em perfis de dispositivo, conteúdo, usuário e serviço de rede”. *Anais do Simpósio Brasileiro de Redes de Computadores*, vol. 2, pp. 554–568.
- W3C “Ubiquitous Web Domain”. <<http://www.w3.org/UbiWeb/>>
- Wang, X.H. et al. (2004) “Ontology based context modeling and reasoning using OWL”. *Anais da IEEE Conference on Pervasive Computing and Communications*, pp. 18–22.
- Yao, Y., Su, S. e Yang, F. (2006) “Service Matching Based on Semantic Descriptions”, *Anais da Conference on Internet and Web Applications and Services*, pp. 126-131.