

# **Uma Análise de Metadados para o Acesso Unificado às Informações dos Repositórios de Ferramentas de Software Livre**

LUCIANO TADEU ESTEVES PANSANATO<sup>1,2</sup>  
RENATA PONTIN DE MATTOS FORTES<sup>2</sup>

<sup>1</sup>CENTRO FEDERAL DE EDUCAÇÃO TECNOLÓGICA DO PARANÁ (CEFET-PR)  
Av. Alberto Carazzai, 1640  
86300-000, Cornélio Procópio, PR  
luciano@cp.cefetpr.br

<sup>2</sup>INSTITUTO DE CIÊNCIAS MATEMÁTICAS E DE COMPUTAÇÃO (ICMC-USP)  
Av. do Trabalhador São-Carlense, 400  
13560-970, São Carlos, SP  
renata@icmc.usp.br

## **Resumo**

Neste artigo são descritas diversas categorias de metadados encontradas na literatura e uma análise dos repositórios de ferramentas de software livre, em relação a essas categorias. O objetivo é identificar categorias comuns para o acesso unificado às informações das ferramentas e facilitar a localização de informação específica. Uma arquitetura para esse acesso e para facilitar a localização de informação também é apresentada.

**PALAVRAS-CHAVE:** software livre, metadados, hipertexto/hipermídia.

## **Abstract**

This paper describes several metadata categories found in the literature and an analysis of repositories of free software tools in relation to metadata categories. The aim is to identify common categories for supporting unified access to the information and facilitating specific information seeking. Architecture to this access and to facilitate the information seeking also is presented.

**KEYWORDS:** free software, metadata, hypertext/hypermedia.

# 1 Introdução

O interesse por projetos de software livre em ambiente empresarial e no meio acadêmico tem sido crescente. No entanto, esses mesmos ambientes, que apresentam atraentes novidades, se mostram deficitários no que diz respeito à adoção de atividades de engenharia de software. De fato, há uma lacuna no que diz respeito à integração de ferramentas de projetos de software livre de apoio às atividades de garantia de qualidade de software e, em particular, aos processos de software livre. Com a tendência atual de utilização de software livre, torna-se relevante viabilizar um processo que atenda às necessidades dessa nova abordagem de desenvolvimento e que contribua efetivamente para a obtenção de produtos de alta qualidade.

A disponibilização de um *framework* relacionado ao contexto de processos de software livre é particularmente interessante por viabilizar a combinação de métodos, ferramentas e procedimentos em uma única estrutura. Nesse contexto, o projeto SAFE<sup>1</sup> [1] [2] (*Software Engineering Available For Everyone* – Engenharia de Software Disponível para Todos) propõe o desenvolvimento de uma infra-estrutura para a integração de ferramentas de software livre de apoio às atividades de Engenharia de Software. Esse *framework* possibilitará um suporte automatizado para o processo de software livre, simples o suficiente para atrair a colaboração e a participação de desenvolvedores nos diversos níveis de familiaridade com o processo de software livre.

O papel de cada ferramenta do *framework* é apoiar uma ou mais atividades do processo de software livre, dependendo dos objetivos de cada projeto específico. Em geral, cada ferramenta armazena os produtos intermediários e também as informações importantes sobre esses produtos e sobre o próprio processo. O desenvolvedor trabalha de maneira integrada com outras ferramentas. No processo de desenvolvimento, determinadas informações geradas em uma determinada etapa podem ser necessárias em outras etapas (e isso ocorre freqüentemente durante o desenvolvimento de software).

A necessidade de localizar uma informação específica pode exigir que o desenvolvedor tenha de pensar em ferramentas diferentes e interfaces diferentes, em cada passo da tarefa de localizar a informação desejada, e o esforço cognitivo necessário para cumprir essa tarefa pode ser considerável. Além disso, as *search engines*, ferramentas tradicionais para situações como essa, não consideram o conteúdo desses repositórios, pois o programa de coleta de páginas (*crawler*) não consegue alcançar essas informações. Portanto, é necessária uma maneira de auxiliar o acesso às informações distribuídas nos diversos repositórios das ferramentas e disponibilizar ao usuário uma interface unificada para facilitar a localização de informação específica.

Nesse contexto, a utilização de metadados como uma infra-estrutura para auxiliar o acesso e facilitar a descoberta<sup>2</sup> dessas informações é uma alternativa a ser considerada. Teevan et al. [3] apresentam um estudo sobre como as pessoas procuram por informação dentro do seu próprio espaço de informação. Nesse estudo, os autores identificaram que as pessoas utilizaram uma quantidade considerável de metadados para localizar a informação desejada.

Neste artigo, é apresentada uma análise das informações gerenciadas por diversas ferramentas de software livre, para identificar os metadados que podem ser considerados para auxiliar o acesso a essas informações. Além disso, é apresentada uma discussão de

---

<sup>1</sup> Projeto em desenvolvimento por pesquisadores do Instituto de Ciências Matemáticas e de Computação (ICMC-USP), com parceria do Departamento de Computação e Estatística (DCT-UFMS), e com a colaboração da empresa Async Serviços de Informática.

<sup>2</sup> O termo descoberta (*discovery*) é um termo geral utilizado no sentido de “encontrar conteúdo”, que abrange outros termos, como busca por consulta (*searching*) e busca por navegação (*browsing*) [4].

uma arquitetura para integrar a informação distribuída nas bases de dados das ferramentas e fornecer uma única interface de acesso.

Este artigo está organizado como segue. As ferramentas de software livre, consideradas para análise, são descritas na Seção 2. As categorias de metadados encontradas na literatura e a análise dos repositórios das ferramentas, em relação a essas categorias, na Seção 3. Uma arquitetura para o acesso unificado, por meio de metadados, aos recursos de informação mantidos por essas ferramentas, é discutida na Seção 4. No final, são apresentadas algumas conclusões.

## 2 Ferramentas de software livre

Uma ferramenta licenciada como software livre é aquela que possui código fonte publicamente disponível, e pode ser alterada e redistribuída livremente. Essa característica faz com que as ferramentas de software livre sejam especialmente adequadas a grupos de desenvolvimento que procuram ferramentas com custo acessível, flexibilidade e garantia de sobrevida.

Nesta seção, são descritas as ferramentas de software livre consideradas para a análise da possibilidade de integração das informações por meio de metadados. Essas ferramentas são algumas adotadas no desenvolvimento do Mozilla<sup>3</sup>, e outras desenvolvidas pelos grupos de pesquisa do Instituto de Ciências Matemáticas e de Computação da Universidade de São Paulo (ICMC-USP). Algumas dessas ferramentas estão disponíveis no SourceForge<sup>4</sup>. O objetivo é identificar o principal recurso de informação (os dados) gerenciado no contexto de cada uma dessas ferramentas. As demais informações utilizadas para descrever o conteúdo desses dados e para representar aspectos relacionados ao gerenciamento, preservação, contexto, e histórico de uso, são consideradas os metadados dos recursos de informação dessas ferramentas.

Por razões de espaço, somente uma descrição concisa sobre cada ferramenta é apresentada. Na Subseção 2.1, é descrita a ferramenta Bugzilla. Na Subseção 2.2, é descrita a ferramenta Bonsai. Na Subseção 2.3, é descrita a VersionWeb. Na Subseção 2.4, a No Risk Planning e, na Subseção 2.5, a DocRationale. Por último, na Subseção 2.6, é descrita a Wiki-RE.

### 2.1. Bugzilla

A ferramenta Bugzilla<sup>5</sup> é um sistema de rastreamento de defeitos (ou *bugs*, na terminologia da ferramenta). Essa ferramenta permite o rastreamento de *bugs* e de mudanças de código, a comunicação entre membros de equipe, a submissão e revisão de atualizações (*patches*), e o gerenciamento da garantia de qualidade. Embora a funcionalidade da ferramenta seja extensa, as operações principais oferecidas ao usuário são simples e diretamente ligadas às atividades de manutenção [5]: a criação de um novo *bug*; a triagem, revisão, escalonamento, discussão e fechamento desse *bug*; e a obtenção de relatórios e consultas baseados em propriedades do *bug*. Assim, os recursos de informação gerenciados pela ferramenta são o código e a documentação sobre os *bugs*.

---

<sup>3</sup> <http://www.mozilla.org/>

<sup>4</sup> <http://sourceforge.net/>

<sup>5</sup> <http://www.mozilla.org/bugs/>

## 2.2. Bonsai

A ferramenta Bonsai<sup>6</sup> é uma interface baseada na Web para o sistema de controle de versões CVS<sup>7</sup> (*Concurrent Versions Systems*), utilizado no projeto Mozilla. O CVS armazena todas as informações sobre o controle de versões em arquivos contidos em uma hierarquia de diretórios (denominada de repositório). Algumas características dessa ferramenta são: permite que os usuários consultem os últimos *check-ins* que ocorreram no repositório CVS; fornece uma interface que facilita a visualização de diferenças entre versões de arquivos do repositório CVS; permite a identificação visual de quais desenvolvedores são responsáveis por quais seções de um código. Em resumo, a Bonsai é um sistema que possibilita a realização de uma grande variedade de consultas relacionando as informações armazenadas no repositório CVS. Assim, os recursos de informação dessa ferramenta abrangem todos os artefatos de um produto de software que são desenvolvidos e modificados, incluindo código fonte, documentação, arquivos executáveis, e outros itens.

## 2.3. VersionWeb

A VersionWeb<sup>8</sup> é uma ferramenta que foi originalmente desenvolvida como uma interface web para o CVS [6]. Atualmente, encontra-se em desenvolvimento a extensão dessa ferramenta para que funcione como uma camada extra para o CVS, para permitir o gerenciamento de grupos de usuários, e o controle de permissões de acesso para esses grupos, além de fornecer os mecanismos para as operações comuns sobre os arquivos contidos nos repositórios do sistema [7]. Assim, a VersionWeb abrange os recursos do CVS e, portanto, os mesmos da ferramenta Bonsai, mas adiciona uma camada de metadados para administração e controle desses recursos com suporte a grupos de usuários.

## 2.4. No Risk Planning

A ferramenta No Risk Planning<sup>9</sup> é um sistema baseado na Web para agendar atividades em grupos a partir das informações do grupo e das informações pessoais de seus membros [8], de maneira semelhante a uma agenda. Inicialmente, a agenda No Risk Planning foi desenvolvida em um projeto de graduação [9] e, depois, o processo de engenharia, manutenção, e evolução desse sistema foi tema de uma dissertação de mestrado [10]. É importante ressaltar que essa agenda está inserida em outros trabalhos de pesquisa, e constantemente recebe um acréscimo de funcionalidades. Por exemplo, a agenda No Risk Planning também fornece um sistema de reserva de salas [11], com o objetivo de auxiliar atividades cotidianas referentes à alocação do recurso “salas” de uma instituição de ensino. Essa funcionalidade está centrada no aspecto de *groupware*, pois as informações acerca da disponibilidade de salas devem ser compartilhadas e sincronizadas entre os grupos de usuários, alunos, professores e funcionários. Assim, os recursos gerenciados são as informações relacionadas às atividades de grupos e dos membros desses grupos, além de outros recursos no contexto de apoio ao trabalho cooperativo, por exemplo, documentos e arquivos compartilhados, mensagens (de bate-papo) trocadas entre membros do grupo, e compromissos relativos a disciplinas (por exemplo, provas, aulas, monitorias e laboratórios).

---

<sup>6</sup> <http://www.mozilla.org/projects/bonsai/>

<sup>7</sup> <https://www.cvshome.org/>

<sup>8</sup> <http://versionweb.sourceforge.net>

<sup>9</sup> <http://nrp.sourceforge.net>

## 2.5. DocRationale

A DocRationale [12] é uma ferramenta baseada na Web que permite a captura, estruturação, armazenamento e recuperação de razões de projeto (*Design Rationale*) relacionadas aos artefatos de software. De acordo com Souza et al. [13], *Design Rationale* (DR) é a documentação das decisões de projeto com suas respectivas justificativas, opções consideradas, avaliações e argumentação que levaram a determinada decisão. A ferramenta possui uma representação simples de DR, para facilitar seu uso e não acarretar esforço excessivo aos desenvolvedores durante a entrada de informações. O objetivo da DocRationale não é fazer inferências a respeito de uma base de informações de projetos, mas documentar informações importantes de projeto visando o reuso dessas mesmas informações em outros projetos [12]. Para tanto, a ferramenta prioriza o fator de colaboração e troca de informações entre os membros da equipe de desenvolvimento para a obtenção de DR. Além disso, também faz uso de anotações, consideradas um bom meio de captura de DR [13]. Assim, os recursos gerenciados são o conjunto de documentos que registra o conhecimento, as suposições e as razões das decisões de projeto relacionadas aos artefatos de software.

## 2.6. Wiki-RE

Atualmente, encontra-se em desenvolvimento uma ferramenta do tipo *wiki*<sup>10</sup>, denominada inicialmente Wiki-RE [14], para auxiliar a engenharia de requisitos por meio da construção de um hiperdocumento de requisitos. A proposta desse trabalho é investigar a utilização de hiperdocumentos para apoiar o pessoal envolvido no processo de engenharia de requisitos, na tarefa de armazenar e recuperar os documentos, artefatos, e experiências gerados pelas técnicas empregadas. A próxima etapa será o desenvolvimento da ferramenta de suporte a autoria distribuída para discussão, armazenamento e consolidação dos requisitos. Assim, o recurso gerenciado é o conjunto de documentos (ou hiperdocumentos) que registra os requisitos de um projeto.

## 3 Categorias de metadados

A idéia de utilizar metadados para facilitar a descrição e descoberta de informação (ou de recursos de informação) não é nova. Os padrões de metadados possuem uma base sólida na área de Biblioteconomia, Ciência da Informação e Arquivologia, anterior à Internet e Web. Em geral, essas abordagens baseadas em metadados para localizar informação permitem ao usuário realizar uma busca, por um objeto armazenado em um repositório, utilizando campos de consulta típicos, como título, autor, ou ano de publicação. No entanto, metadados não apenas descrevem o conteúdo de outros dados, mas também indicam muitos outros aspectos, por exemplo, os aspectos relacionados ao gerenciamento, preservação, contexto, e histórico de uso de recursos de informação [15]. Isso motivou a realização de um levantamento das categorias de metadados, para servir de base para uma análise em alto nível dos metadados contidos nos repositórios das diversas ferramentas de software livre, com o objetivo de verificar a possibilidade da utilização de metadados para o acesso unificado às informações desses repositórios.

Algumas classificações de metadados são encontradas na literatura [16] [17] [18] [15]. Em resumo, a classificação de Moura et al. [18] abrange as classificações de Böhm e Rakow [16] e Kashyap et al. [17], e foi estendida com as categorias propostas por Gilliland-

---

<sup>10</sup> Um website colaborativo que inclui o trabalho coletivo e constante de diversos autores (<http://www.webopedia.com/TERM/W/wiki.html>).

Swetland [15]. As categorias, divididas em quatro grupos funcionais, são descritas nas subseções seguintes. Na Subseção 3.1, as categorias de metadados para descoberta de recursos são descritas. Na Subseção 3.2, as categorias para disponibilidade de recursos são descritas. Na Subseção 3.3, as categorias para uso de recursos são descritas. Na Subseção 3.4, são descritas as categorias para administração e controle de recursos. Finalmente, na Subseção 3.5, é apresentada uma análise das informações dos repositórios das ferramentas descritas na Seção 2, segundo as categorias descritas nesta seção.

### 3.1. Categorias de metadados para descoberta de recursos

Os elementos de metadados das categorias para descoberta de recursos representam o conjunto de termos necessários para descobrir e identificar recursos de informação. Essas categorias são as seguintes:

- **Metadados para descrição de bibliografia.** Esses metadados são elementos descritivos para a identificação e caracterização do recurso.
- **Metadados para identificação única dos recursos.** Esses metadados fornecem uma identificação única para o recurso.
- **Metadados para descrição geral do conteúdo dos recursos.** Esses metadados correspondem a elementos para descrever o conteúdo do recurso.
- **Metadados para descrição do assunto dos recursos.** Esses metadados podem incluir termos controlados ou não controlados, utilizados para caracterizar os conteúdos de informação do recurso.
- **Metadados para descrição da estrutura dos recursos.** Esses metadados são utilizados para descrever a composição de um recurso.
- **Metadados para descrição do relacionamento entre recursos.** Esses metadados incluem elementos para associar um recurso específico com outros recursos relacionados.
- **Metadados para descrição da proveniência dos recursos.** Esses metadados identificam as principais fontes ou fornecedores do recurso.
- **Metadados para descrição do formato e mídia dos recursos.** Esses metadados correspondem a elementos para a representação dos dados e das características de apresentação dos componentes do recurso.

### 3.2. Categorias de metadados para disponibilidade de recursos

Os elementos de metadados das categorias para disponibilidade de recursos definem os termos e condições necessárias para o acesso e recuperação de recursos, sem considerar se a maneira utilizada é restrita ou irrestrita. Essas categorias são as seguintes:

- **Metadados para distribuição dos recursos.** Esses metadados descrevem como o recurso é disponibilizado.
- **Termos e condições para acesso aos recursos.** Esses metadados especificam as condições prévias de acesso que devem ser preenchidas para garantir a proteção da privacidade e propriedade intelectual associadas ao recurso.
- **Termos e condições para uso dos recursos.** Esses metadados descrevem os termos e condições para utilizar o recurso.
- **Metadados sobre requisitos dos recursos.** Esses metadados especificam os requisitos de software e hardware para o uso do recurso.

- **Metadados para localização dos recursos.** Esses metadados fornecem a informação necessária para transferência do recurso. O principal objetivo é permitir ao sistema selecionar uma cópia ou versão adequada do recurso (por exemplo, para uso local).
- **Metadados sobre autenticidade dos recursos.** Esses metadados descrevem esquemas ou métodos para garantir a autenticidade de um recurso.

### 3.3. Categorias de metadados para uso de recursos

Os elementos de metadados das categorias para uso do recurso são informações adicionais para permitir o uso adequado de recursos. Essas categorias são as seguintes:

- **Metadados para classificação do conteúdo dos recursos.** Esses metadados representam classificações, baseadas no conteúdo do recurso, de acordo com algum esquema contextual atribuído por uma autoridade.
- **Metadados para descrição da qualidade dos recursos.** Esses metadados especificam a qualidade dos dados do recurso.
- **Metadados para descrição do objetivo dos recursos.** Esses metadados descrevem porque o recurso de informação é fornecido e identifica os programas, projetos, fóruns de discussões, etc., relacionados a esse recurso.
- **Metadados para descrição contextual dos recursos.** Esses metadados fornecem informação relacionada a eventos específicos, situações, configurações, etc., relacionados ao domínio ou finalidade do recurso.

### 3.4. Categorias de metadados para administração e controle de recursos

Os elementos de metadados das categorias para administração e controle de recursos fornecem informação para controlar, auditar e rastrear a informação sobre o próprio recurso e também sobre os seus metadados. Essas categorias são as seguintes:

- **Metadados para controle de modificação dos recursos.** Esses metadados controlam as versões do recurso de informação.
- **Metadados para administração dos recursos.** Esses metadados são relacionados a qualquer informação com relação ao gerenciamento e controle do próprio recurso de informação.
- **Metadados para histórico de uso dos recursos.** Esses metadados são reservados para representar informação sobre as operações executadas no recurso de informação, como cópia, edição, remoção, etc.
- **Metadados para administração dos metadados dos recursos.** Esses metadados são relacionados a qualquer informação com relação ao gerenciamento e controle dos próprios metadados do recurso de informação.
- **Metadados para preservação dos recursos.** Esses metadados são relacionados ao gerenciamento da preservação do recurso de informação.

### 3.5. Análise

Nesta subseção, as ferramentas são comparadas de acordo com a cobertura dos seus elementos de metadados, utilizando as categorias descritas anteriormente. Para tanto, considera-se como metadados as informações utilizadas para descrever o conteúdo e os aspectos relacionados ao gerenciamento, preservação, contexto e histórico de uso do

principal recurso de informação de cada ferramenta, identificado na Seção 2. Para a realização da análise desses metadados, foi necessária a leitura cuidadosa das referências e da documentação (manuais e tutoriais) das ferramentas, a instalação e uso das ferramentas, e a verificação detalhada dos repositórios de informações (base de dados). Um resumo dessa comparação é apresentado no Quadro 1.

Quadro 1 – Comparação entre ferramentas de software livre, segundo a cobertura dos seus elementos de metadados.

Conjunto de Elementos de Metadados		Bugzilla	Bonsai	VersionWeb	No Risk Planning	DocRationale	Wiki-RE
Categorias de Metadados para descoberta de recursos	Metadados para descrição de bibliografia						
	Metadados para identificação única dos recursos	X	X	X	X	X	X
	Metadados para descrição geral do conteúdo dos recursos	X	X	X	X	X	X
	Metadados para descrição do assunto dos recursos						
	Metadados para descrição da estrutura dos recursos		X	X		X	X
	Metadados para descrição do relacionamento entre recursos	X	X	X	X	X	X
	Metadados para descrição da proveniência dos recursos	X	X	X	X	X	X
	Metadados para descrição do formato e mídia dos recursos						
Metadados para disponibilidade de recursos	Metadados para distribuição dos recursos						
	Termos e condições para acesso aos recursos						
	Termos e condições para uso dos recursos						
	Metadados sobre requisitos dos recursos		X	X			
	Metadados para localização dos recursos		X	X			
	Metadados sobre autenticidade dos recursos						
Metadados para uso de recursos	Metadados para classificação do conteúdo dos recursos	X	X	X	X	X	X
	Metadados para descrever qualidade dos recursos						
	Metadados para descrição do objetivo dos recursos						
	Metadados para descrição contextual dos recursos	X					X
Metadados para administração e controle de recursos	Metadados para controle de modificação dos recursos	X	X	X			X
	Metadados para administração dos recursos	X		X	X	X	X
	Metadados para histórico de uso dos recursos	X	X	X	X	X	X
	Metadados para administração dos metadados dos recursos						
	Metadados para preservação dos recursos		X	X			

Nos resultados dessa comparação, observa-se que as ferramentas possuem diversas categorias comuns de metadados, as quais indicam a possibilidade de utilizar esses metadados para auxiliar o acesso unificado às informações distribuídas pelos diferentes repositórios das ferramentas. Outro resultado importante é a identificação de quais categorias de metadados são negligenciadas e cuja inclusão deve ser considerada (estudo de viabilidade) nas próximas versões das ferramentas.

Para obter a interoperabilidade semântica desejada para o acesso unificado, é necessário realizar uma análise e correlação dos elementos de dados dos repositórios, para obter um mapeamento dos metadados que possuem significado igual ou similar. Nos trabalhos iniciais, com alguns dos repositórios, nota-se que essa tarefa é dificultada pelas diferenças dos conteúdos dos elementos de dados. Em geral, essas diferenças são devido ao próprio formato do conteúdo (por exemplo, elemento estruturado e elemento textual com o mesmo significado) e aos vocabulários diferentes adotados pelas ferramentas. Além disso, em alguns casos, os elementos considerados relevantes não possuem correspondente em



determinados repositórios ou não é possível conseguir um mapeamento em ambos os sentidos.

#### 4 Acesso unificado aos recursos de informação de diversas ferramentas de software livre por meio de metadados

Nesta seção, é apresentada uma arquitetura, em desenvolvimento por pesquisadores do ICMC-USP, para acesso unificado aos repositórios de informações de ferramentas de software livre, como mostrada na Figura 1. Essa arquitetura utiliza metadados como infraestrutura para auxiliar a localização de informação relacionada com os metadados do conteúdo da página de uma ferramenta. Em alto nível, a arquitetura consiste do navegador, de um servidor com os componentes principais, e das ferramentas de software livre estendidas por meio de um componente para auxiliar o acesso aos repositórios dessas ferramentas. A seguir, é apresentada uma breve discussão sobre cada componente dessa arquitetura:

- **Componente de gerência de acesso.** Esse componente, implementado como um *Web proxy*, é responsável pelo gerenciamento do acesso aos repositórios das ferramentas de software livre. A idéia básica é introduzir um *Web proxy* como uma camada intermediária entre o navegador e as ferramentas. Dessa maneira, o acesso unificado é transparente para o usuário, isto é, o usuário permanece alheio a qualquer detalhe relacionado aos repositórios distribuídos. O usuário apenas configura o *proxy* no navegador e, adicionalmente, pode estabelecer suas preferências em relação à interface e ao comportamento da integração, por exemplo, quais ferramentas são consultadas para localizar informação relacionada. Diversos *proxies* podem estar disponíveis em máquinas diferentes para evitar a dependência de um único serviço.
- **Componente de autenticação e preferências de usuário.** Esse componente é responsável pelo controle de autenticação e pela recuperação das preferências de usuário. São exemplos de preferências de usuário: detalhes de interface de navegação, conjunto de metadados para filtragem e/ou agrupamento, conjunto de ferramentas consultadas pelo componente de gerência de acesso.
- **Componente de extração de metadados.** Esse componente é um aspecto chave da arquitetura e consiste de um programa que extrai os metadados da página solicitada pelo usuário segundo uma estratégia (heurística) definida para cada ferramenta. A abordagem utilizada é baseada na extração de conteúdo de documentos HTML utilizando DOM (*Document Object Model*), semelhante a [19].
- **Componente de acesso.** Esse componente, implementado como um *Web service* [20], é responsável por localizar informação relacionada no repositório da ferramenta, baseado em um conjunto de metadados, e enviar uma lista de links e informações de contexto ao componente de gerência de acesso. O acesso é facilitado pela utilização de *Web services* porque cada ferramenta tem a sua própria maneira de gerenciamento dos seus recursos de informação. Além disso, a inclusão de novas ferramentas também é facilitada, pois somente é necessário que a nova ferramenta disponibilize um *Web service*. No *Web service* são implementadas as estratégias de mapeamento dos metadados e as consultas necessárias para localização de informação no repositório mantido pela ferramenta. É importante destacar que não é necessária a criação de índices, pois o *Web service* pode manipular diretamente o repositório da ferramenta.
- **Componente de geração de página.** Esse componente é responsável por combinar as listas (de links e informações de contexto) fornecidas pelos componentes de acesso e incluir uma única lista na página da ferramenta solicitada pelo usuário, para fornecer

acesso direto às informações relacionadas. No entanto, a lista final de informações relacionadas pode ser extensa. Para minimizar a sobrecarga de informação, o componente de geração de página também é responsável por incorporar um suporte à navegação para facilitar a localização de informação específica no conjunto de informações resultante das consultas aos componentes de acesso (*Web services*) das ferramentas. Esse suporte fornece diversos mecanismos baseados em metadados para localizar a informação desejada: busca por palavra-chave, busca por navegação, filtragem e agrupamento, entre outros.

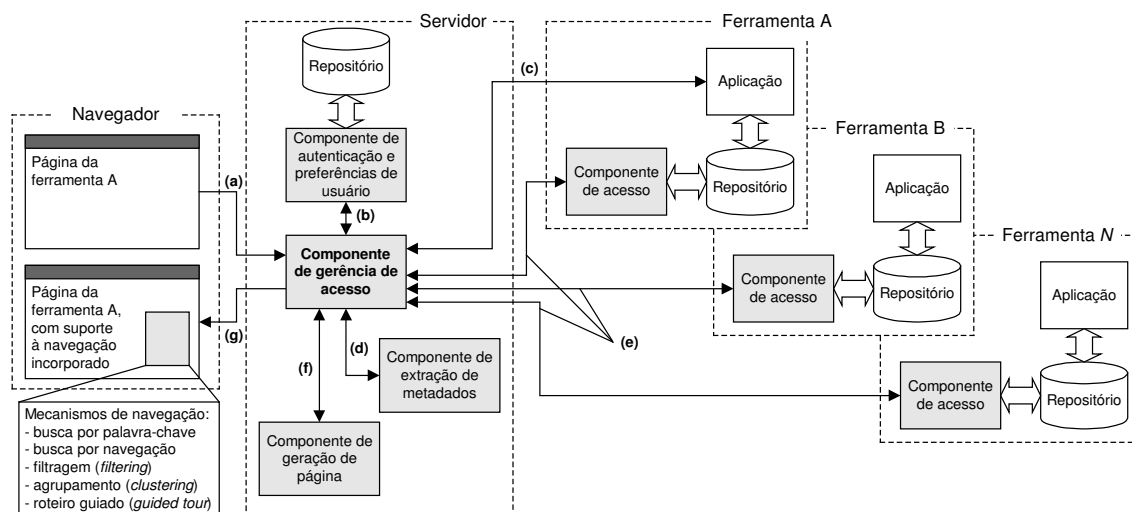


Figura 1 – A arquitetura de acesso unificado: (a) comunicação do usuário, por meio do navegador, com o componente de gestão de acesso (*Web proxy*), para solicitar uma nova página da ferramenta A; (b) controle de autenticação e recuperação das preferências de usuário, (c) comunicação do componente de gestão de acesso com a ferramenta A, para executar a solicitação do usuário; (d) extração dos metadados da nova página da ferramenta A; (e) comunicação do componente de gestão de acesso com os componentes de acesso (*Web services*), para solicitação das páginas com informações relacionadas aos metadados extraídos; (f) comunicação do componente de gestão de integração com o componente de geração de página, para incorporar uma lista de links (para páginas das outras ferramentas relacionadas com a nova página da ferramenta A) e mecanismos de navegação; (g) comunicação do componente de gestão de integração com o usuário, por meio do navegador.

Na abordagem adotada para a arquitetura, o servidor (componente de gestão de acesso) deve encaminhar a solicitação para a ferramenta, receber a página de resposta e fazer a extração dos metadados (componente de extração de metadados), estabelecer múltiplas sessões de consultas com cada repositório (componente de acesso), combinar os resultados e gerar a página final para apresentação ao usuário (componente de geração de página). O processo completo tem um custo computacional considerável que pode afetar o tempo de resposta para o usuário. Além disso, os componentes de acesso são específicos para cada repositório das ferramentas, e isso dificulta a automatização do processo de implementação desses componentes.

## 5 Conclusões

A existência de diversas categorias de metadados comuns a todas as ferramentas de software livre consideradas na análise, indica a possibilidade de utilizar esses metadados para facilitar a localização de informação específica nos diversos repositórios dessas

ferramentas. A facilidade com a qual uma informação pode ser encontrada depende da estratégia utilizada para manipular os metadados de maneira a garantir a inclusão da informação desejada em uma lista pequena o suficiente para o usuário encontrar a informação. Na arquitetura apresentada, essa questão é abordada por meio do acesso unificado às informações distribuídas nos diversos repositórios das ferramentas, utilizando metadados como infra-estrutura. Os metadados contidos na página solicitada pelo usuário para uma ferramenta são extraídos e utilizados para recuperar, dos repositórios das outras ferramentas, as informações relacionadas. A arquitetura fornece uma interface unificada, a partir da qual, o acesso à informação relacionada à página atual é obtido diretamente. Assim, esse mecanismo simples pode antecipar a necessidade do usuário, facilitando a localização da informação e até mesmo a descoberta de informação que, de outra maneira, poderia não ser considerada.

## Agradecimentos

Agradecemos a FINEP, pelo patrocínio ao projeto, e à agência CAPES pelo apoio financeiro aos autores.

## Referências

- [1] Fortes, R. P. M., Turine, M. A. S., Reis, C. R. *Engenharia de Software Disponível a Todos - Software Engineering Available for Everyone (SAFE)*. Proposta de Projeto para Chamada CT-Info-FINEP, 2004.
- [2] Fortes, R. P. M., Silva, M. A. G., Freire, A. P., Junqueira, D. C. SAFE - Software Engineering Available For Everyone. In *Anais do V Workshop sobre Software Livre*, Porto Alegre, 2004, p. 203-206.
- [3] Teevan, T., Alvarado, C., Ackerman, M. S., Karger, D. R. The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *Proceedings of the 2004 Conference on Human Factors in Computing Systems*, Vienna, Austria, 2004, p. 415-422.
- [4] W3C, World Wide Web Consortium. *RDF Primer*, W3C Recommendation, 2004. <http://www.w3.org/TR/rdf-primer/>, acesso em 1° de junho de 2004.
- [5] Reis, C. R. Uma Visão Geral do Bugzilla, uma Ferramenta de Acompanhamento de Alterações. In *Sessão de Ferramentas do SBES'2002 - 16º Simpósio Brasileiro de Engenharia de Software*, Gramado, 2002.
- [6] Moreira, D. A., Soares, M., Fortes, R. P. M. VersionWeb: A tool for helping web page version control. *Linux Journal*, 2002. <http://www.linuxjournal.com/article.php?sid=6330>, acesso em 2 de julho de 2004.
- [7] Fortes, R. P. M., Freire, A. P., Vieira, V. H., Paiva, D. M. B. An Academic Web-Based Agenda and its Engineering Process. In *VII Workshop Iberoamericano de Ingeniería y Desarrollo de Ambientes de Software - IDEAS*, Arequipa, Peru, 2004, p. 151-156.
- [8] Junqueira, D. C., Fortes, R. P. M. Considerações sobre controle de grupos em repositórios CVS durante a evolução de uma ferramenta Web. In *Anais do V Workshop sobre Software Livre*. Porto Alegre, 2004, p. 129-132.
- [9] Moura, M. V. *No Risk Planning - um sistema de suporte ao trabalho em grupo: uma proposta para Web*. São Carlos, 2001. Monografia de Projeto de Graduação. Instituto de Ciências Matemáticas e de Computação de São Carlos, Universidade de São Paulo.
- [10] Ribeiro, T. M. *Evolução de uma agenda para grupos na Web*. São Carlos, SP, 2003, Dissertação (Mestrado). Instituto de Ciências Matemáticas e de Computação de São Carlos, Universidade de São Paulo.
- [11] Freire, A.P., Fortes, R.P.M. *Documentação do Sistema de Reserva de Salas da Intranet do ICMC-USP*, São Carlos, SP, Instituto de Ciências Matemáticas e de Computação de São Carlos, Universidade de São Paulo, 2003. 48p. (Relatórios Técnicos do ICMC, 222).
- [12] Francisco, S.D. *DocRationale - uma ferramenta para suporte a Design Rationale de artefatos de Software*. São Carlos, SP, 2004. 123p. Dissertação (Mestrado). Instituto de Ciências Matemáticas e de Computação de São Carlos, Universidade de São Paulo.

- [13] Souza, C. R. B., Wainer, J., Santos, D. B., Dias, K. L. A model and tool for semi-automatic recording of design rationale in software diagrams. In *Proceedings of the 6th String Processing and Information Retrieval Symposium & 5th International Workshop on Groupware*, Cancun, Mexico, 1998, p. 306–313.
- [14] Silva, M. A. G., Fortes, R. P. M. Processo de Engenharia de Requisitos com Apoio de Hiperdocumentos. In *Workshop de Teses e Dissertações em Engenharia de Software*, Manaus, 2003, p. 65–70.
- [15] Gilliland-Swetland, A. Setting the Stage, In *Introduction to Metadata: Pathways to Digital Information*, T. Gill, A. Gilliland-Swetland, M. Baca, Getty Education Institute for the Arts, Los Angeles, CA, <http://www.getty.edu/research/institute/standards/intrometadata/>, Acesso em: 1º de junho de 2004.
- [16] Bohn, K, Rakow, T. C. Metadata for Multimedia Documents. *SIGMOD Record*, 23(4), 1994, p. 21-26
- [17] Kashyap, V., Shah, K., Sheth, A. Metadata for Building the Multimedia Path Quilt. In *Multimedia Database System: Issues and Research Directions*, Springer-Verlag, 1995.
- [18] Moura, A. M. C., Campos, M. L. M., Barreto, C. M. A Survey on Metadata for Describing and Retrieving Internet Resources. *World Wide Web Journal*, 1(4), 1998, p. 221-240.
- [19] Gupta, S., Kaiser, G., Neistadt, D., Grimm, P. DOM-based Content Extraction of HTML Documents. In *Proceedings of 12th International Conference on World Wide Web*, Budapest, Hungary, 2003.
- [20] Stal, M. Web Services: beyond component-based computing. *Communications of the ACM*, 45(10), 2002, p. 71–76.