Software Livre na Informatização de Acervos Históricos

Elisa Yumi Nakagawa[1*]

[1]Depto de Ciências Administrativas e Tecnologia
Centro Universitário de Araraquara - Uniara
Rua Voluntários da Pátria, 1309
14801-320, Araraquara, SP

elisa@icmc.usp.br

Abstract. *Nowadays, the brazilian historical holding, including those in São Carlos and region, are sometimes unknown or unavailable. Therefore, a free Web system that becomes possible the storage and dissemination of information about these holdings is extremely relevant. Thus, it is necessary the identification of an attributes set based on international norm and register format to represent the information about several types of historical holdings. Consequently, in the next future, the system can propitiate the implantation of a distributed and integrated database of the several historical holdings and avail it to all community.*

Resumo. *Atualmente, os acervos históricos brasileiros, inclusive os da cidade de São Carlos e região, encontram-se, muitas vezes, desconhecidos ou indisponíveis; dessa forma, um sistema Web livre que possibilite o armazenamento e disponibilização de informações desses acervos é de extrema relevância. Para isso, é indispensável a identificação de um conjunto de atributos baseado em normas e formatos de registro internacionais que possam representar integralmente as informações sobre os diversos tipos de acervos históricos. Como conseqüência e sendo um software livre, o sistema poderá propiciar em um futuro próximo, a implantação de uma base de dados distribuída e integrada dos diversos acervos históricos disponibilizada a toda a comunidade.*

1. Introdução

O município de São Carlos e região apresenta uma diversidade de acervos de inestimável valor histórico. Esses acervos contêm elementos de caráter diversificado de extrema relevância para o resgate da história do município e inclusive relacionados à própria história do Brasil. Atualmente, esses acervos, bem como grande parte dos acervos brasileiros, encontram-se, muitas vezes, em estado precário de conservação e outros ainda, desconhecidos até mesmo por pesquisadores da área. Verifica-se nos últimos dois anos diversos esforços isolados, tanto público quanto privado, no sentido de organizar os acervos e disponibilizá-los aos interessados.

O projeto Memória Virtual de São Carlos [Mem, 2003] agrega as competências estabelecidas no meio acadêmico[1] a serviço de uma iniciativa bastante relevante para a comunidade da cidade de São Carlos e região. O projeto, de natureza multi-disciplinar e envolvendo a Prefeitura Municipal de São Carlos, a Fundação Pró-Memória de São Carlos, a Associação Pró-Casa do Pinhal e outras instituições, pode contribuir de forma significativa para o registro histórico de São Carlos, bem como para a ampliação do acesso a esse registro. Para isso, o principal objetivo desse projeto é o desenvolvimento de um sistema Web que será disponibilizado como software

[*]É também professora e doutoranda do programa de pós-graduação do Instituto de Ciências Matemáticas e de Computação - ICMC/USP.

[1]Participam desse projeto pesquisadores do Instituto de Ciências Matemáticas e de Computação - ICMC/USP, do Departamento de Arquitetura e Urbanismo - SAP/USP, do Departamento de Ciência da Informação - DCI/UFSCar, da Biblioteca Comunitária - BCO/UFSCar e do Instituto de Biociências, Letras e Ciências Exatas - IBILCE/UNESP.

livre[2] para a gestão integrada de acervos do município; mais especificamente, para organização, armazenamento e disponibilização de informações do patrimônio histórico e cultural, possibilitando a reunião dos diversos acervos dispersos fisicamente, seguindo normas nacionais e internacionais de definição de conjunto de atributos para representação da informação. Esse sistema possibilitará também promover a conscientização sobre a importância de preservar a grande diversidade de acervos históricos. Vale ressaltar que, assim como as inúmeras iniciativas de diversos órgãos públicos em nível nacional para uso de software livre, a Prefeitura Municipal de São Carlos também está empenhada na adoção, utilização efetiva e disseminação de software livre; além de estar apoiando fortemente para a concretização desse projeto.

O objetivo deste artigo é então divulgar o Projeto Memória Virtual de São Carlos, sua relevância e principais contribuições, bem como seu relacionamento com o tema software livre. Dessa forma, este artigo está organizado da seguinte forma: na Seção 2 será discutido o desenvolvimento do sistema Web, mais especificamente, as atividades no tocante ao levantamento de requisitos, bem como a identificação de um ambiente de implementação do sistema. E por fim, na Seção 3, encontram-se as conclusões.

2. Desenvolvimento do Sistema Web

2.1. Levantamento de Requisitos

Uma vez que o sistema Web proposto nesse projeto irá gerenciar uma diversidade e um grande volume de informações, o seu planejamento passa a ter um papel estratégico para o acesso pleno a essa informação. O objetivo do projeto tem então demandado pesquisas nas diversas áreas de conhecimento envolvidas. Como um dos exemplos mais sobressalentes é a pesquisa em Ciência da Informação que tem definido um conjunto de atributos tanto para representação, armazenamento e disponibilização de informações dos diversos tipos de acervos históricos, quanto como ponto de acesso às informações. Os resultados dessa pesquisa, inédito em nível internacional, é de extrema relevância como pré-requisito fundamental para a implementação do sistema, mais especificamente para o projeto da base de dados.

Vale ressaltar que a escolha de atributos corretos garante rapidez e precisão na recuperação das informações da base de dados, e conseqüentemente, credibilidade ao sistema. Entende-se por pontos de acesso todos os elementos que descrevem a informação e, concomitantemente, servem como chave para a recuperação da informação e para o seu acesso na íntegra. Os atributos, sendo ou não pontos de acesso, são representações da informação dentro do sistema, e por isso, também são conhecidos como metadados, ou seja, dados sobre um outro dado [Hillmannn, 2004]. Assim, atributos como autoria e palavra-chave, por exemplo, que representam a informação original e/ou o seu conteúdo e permitem a sua recuperação, são metadados dessa informação.

De modo geral, o sistema inclui os seguintes módulos ou conjuntos de informação a serem manipulados: acervos documentais, coleções bibliográficas, objetos museológicos, bens arquitetônico, bens naturais, além de materiais multimeios. Atualmente, está-se concentrado no levantamento de requisitos, projeto e implementação dos três primeiros módulos. Definir um conjunto de atributos para esses conjuntos de informação não é uma tarefa trivial, uma vez que não existe um conjunto já estabelecido que possa servir de base e que contemple as especificidades dos vários conjuntos de informação que serão manipulados pelo sistema proposto. Pode-se identificar conjuntos de atributos[3] — também chamados de formatos de registro — internacionais, consolidados e amplamente divulgados e utilizados para conjuntos específicos de informação. No contexto desse projeto diversos formatos têm sido considerados, dentre eles o MARC (*Machine Readable Cataloguing*) como formato de registro para informações bibliográficas, a EAD (*Encoded Archival Description*) para os arquivos, considerados no todo ou documento por documento, e o Dublin Core para os recursos eletrônicos.

[2] http://www.fsf.org/philosophy/free-sw.html
[3] http://www.ifla.org/II/metadata.htm

Além dos formatos de registro, pode-se identificar também normas internacionais que estabelecem padrões de como descrever a informação, ou seja, como os dados devem ser inseridos na base de dados. Certamente, essas normas são também instrumentos essenciais para a definição do conjunto de atributos, sendo o AACR-2 (*Anglo-American Cataloguing Rules, 2nd. Edition*)[4] e o ISAD(G) (*General International Standard Archival Description*)[5] padrões que têm sido utilizados no âmbito desse projeto. Portanto, a utilização de formatos de registro e de normas internacionais para a descrição da informação devem ser complementares. Com isso, é possível garantir a qualidade das informações armazenadas e disponibilizadas, bem como a possibilidade de integração com outros sistemas de informação, tanto nacionais quanto internacionais.

Além disso, uma vez que sistemas como o almejado não são largamente conhecidos, *brainstorms* têm sido realizados em reuniões entre os membros do projeto para a identificação e definição de requisitos do sistema. Além disso, meios de comunicação eletrônicos, tais como correio eletrônico e sistema de mensagem instantânea, também têm contribuído em muito para a realização dessa atividade. Os requisitos identificados e considerados mais relevantes, são descritos textualmente — baseando-se nas práticas recomendadas pela IEEE para especificação de requisitos do software [IEEE Std 830, 1998] — sendo uma forma de registrar para posterior verificação, bem como para facilitar a integração de novos desenvolvedores, em se tratando de um sistema que possui uma grande quantidade de funcionalidades não comumente conhecidas pela maioria dos potenciais desenvolvedores.

2.2. Projeto e Implementação do Sistema

Para a construção do sistema, está sendo investigada a utilização de práticas do processo de desenvolvimento de software livre — tais como a forma como os requisitos são identificados e a atribuição de tarefas aos desenvolvedores — como descrito inicialmente em "*The Cathedral and the Bazaar*" por [Raymond, 1999] e na mesma linha de diversas pesquisas, sendo algumas delas do ICMC/USP [Nakagawa, 2002, Reis, 2003]. Como abordagem para a implementação do sistema, a prototipação está sendo empregada, como ocorre na grande maioria dos softwares livres sendo desenvolvidos; ou seja, à medida que o sistema é implementado, novos requisitos são identificados e já implementados e, no caso desse projeto, registrados também em um documento de requisitos.

Quanto ao projeto do sistema, técnicas para modelagem da funcionalidade do sistema não estão sendo utilizadas; por outro lado, modelos da base de dados são construídos utilizando DER (Diagrama de Entidade-Relacionamento), como resultado do conjunto de atributos identificado, e também como uma forma de documentar o sistema, uma vez que se trata de uma base de dados de razoável complexidade em termos do número de atributos, relacionamentos e tabelas. Questões relacionadas à usabilidade do sistema em virtude da diversidade de usuários, bem como questões no tocante à segurança e robustez tem sido consideradas em se tratando do volume de dados a serem manipulados. Por exemplo, um dos módulos refere-se às coleções bibliográficas que irá inicialmente gerenciar dados da Biblioteca da Fazenda Pinhal[6] que possui aproximadamente 5000 obras desde livros, cartas, mapas, entre outros.

Sendo o software livre um dos requisitos do ambiente de implementação e implantação do sistema, foi realizado um estudo e seleção das ferramentas/tecnologias a serem utilizadas. Adotou-se uma plataforma de software livre, no caso o Linux[7] e o Apache[8]; como SGBD (Sistema de Gerenciamento de Banco de Dados) adotou-se o PostgreSQL[9] e para a implementação das funcionalidades do sistema está utilizando-se o PHP[10], todos em consonância com a filosofia adotada no projeto. Todas possuem uma vasta comunidade de usuário e uma experiência consagrada na construção de uma diversidade de sistemas dos mais variados domínios de aplicação.

[4] http://www.collectionscanada.ca/jsc/index.html
[5] http://www.hmc.gov.uk/icacds/eng/home.htm
[6] http://www.fazendapinhal.com.br
[7] http://www.linux.org/
[8] http://httpd.apache.org/
[9] http://www.postgresql.org
[10] http://www.php.net/

3. Conclusão

Atualmente, a grande maioria dos acervos históricos brasileiros necessitam de tratamento adequado, bem como uma divulgação no sentido de conscientizar a comunidade de seu valor histórico. Para isso, a tecnologia de computação possui mecanismos que quando utilizados adequadamente possibilitam ser um veículo dos mais relevantes na realização dessa tarefa.

Vale ressaltar que por meio desse projeto, tem ocorrido a constante e enriquecedora troca de experiências entre as diversas áreas de pesquisa envolvidas, o que vem salientar a relevância da realização de projetos multi-disciplinares. No mais, esse projeto tem também contribuído no sentido de identificar a problemática das áreas envolvidas e que necessita ser sanada resultando em pesquisas de valioso teor e fundamental para a concretização dos objetivos propostos no projeto.

Como benefício mais expressivo e imediato do projeto têm se a disponibilização de um conjunto de atributos capaz de armazenar e disponibilizar de forma efetiva informações dos mais diversos tipos de acervos históricos. Além disso, a implementação desse conjunto em um sistema Web livre certamente irá contribuir, tanto para a comunidade de software livre com um sistema ainda não encontrado nesse domínio de aplicação, quanto no sentido de facilitar a implantação e posterior disponibilização de informações sobre os mais diversos acervos para a comunidade em geral. Observa-se que um sistema como o proposto não é encontrado nem como software livre, nem como software proprietário; aqueles identificados possuem funcionalidades ou bases de dados limitadas, o que não está em consonância com os objetivos do projeto; isso vem reforçar a relevância da disponibilização de sistema desse domínio de aplicação como software livre.

Partindo do sistema piloto a ser implantado inicialmente nos acervos pertencentes à rota do café na qual São Carlos está inserido, pretende-se instalar um mecanismo para disseminação desse sistema para diversas outras entidades similares — prefeituras, fazendas históricas, associações, entre outras — a nível regional, estadual e até nacional, construindo-se assim uma base de dados distribuída, no entanto, integrada dos diversos acervos históricos brasileiros dispersos fisicamente.

Agradecimentos: Agradeço a todos os membros do Projeto Memória Virtual de São Carlos e em especial ao Prof. Dr. José Carlos Maldonado e à Juliana de Souza Moraes pelo suporte na concepção deste artigo.

Referências

(2003). Memória Virtual de São Carlos. Instituto de Ciências Matemáticas e de Computação - ICMC/USP, São Carlos, SP. (Projeto aprovado no programa de políticas públicas da FAPESP e coordenado pelo Prof. Dr. José Carlos Maldonado).

Hillmannn, D. (2004). What is metadata? [*On-line*], *World Wide Web*. Disponível em `http://dublincore.org/documents/2001/04/12/usageguide/#whatismetadata` (Acessado em 30/04/2004).

IEEE Std 830 (1998). IEEE recommended practice for software requirements specifications.

Nakagawa, E. Y. (2002). *Software Livre: Processo e Produto Livres no Desenvolvimento de Aplicações Web*. Exame geral de qualificação, Instituto de Ciências Matemáticas e de Computação – ICMC/USP, São Carlos, SP. (Doutorado em andamento).

Raymond, E. S. (1999). *The Cathedral and the Bazaar*. O'Reilly & Associates.

Reis, C. R. (2003). Caracterização de um processo de software para projetos de software livre. Master's thesis, ICMC/USP, São Carlos - SP. Dissertação de Mestrado.