Arquivos Abertos, a OAI e o OAI-PMH

Open Archives Initiative

A Open Archives Initiative é uma organização não governamental cujo objetivo é desenvolver e promover soluções de interoperabilidade, de tal forma a **facilitar o acesso eficiente aos conteúdos em formato digital e na Web**.

A OAI começou com a comunidade acadêmica e sua "literatura cinzenta" (*gray literature*):

- A comunidade acadêmica escrevia seus trabalhos e os publicava em sites na Internet (*e-prints*)
- Os trabalhos eram "arquivados" (perdoem os arquivistas!!) nos discos rígidos dos computadores
- Os computadores podiam (e eram) acessados através da Internet

Existia na Web um **rico conteúdo**(*) de resultados de educação, pesquisa e desenvolvimento (produtos das atividades da comunidade acadêmica) – todo **interconectado através da rede**!!

(*) Este tipo de conteúdo é designado *Scholarly Information* na bibliografia da área.

Com o objetivo de organizar a situação dos *e-prints* e sua disseminação, um grupo de pessoas da área de TIC, da academia e da comunidade de informação, reuniu-se em Santa Fé, NM, em out.1999. Resultados da reunião:

- Criada a OAI
- Iniciado um trabalho para a integração (*federation*) dos conteúdos (*e-prints*) existentes na Web

Passos seguintes:

- Trabalho técnico e administrativo para dar prosseguimento à Iniciativa
- Ampliação do escopo da Iniciativa para todos os conteúdos digitais de natureza acadêmica, transcendendo os *e-prints*
- Consolidação da solução técnica

Duas palavras são importantes no contexto:

- *Open* (abertos) – significa abertura do ponto de vista de arquitetura do sistema, ou seja, de como os máquinas se comunicam; não quer dizer uso gratuito e indiscriminado (as máquinas podem se comunicar e chegar até um conteúdo e constatar que ele tem um preço ou condição de uso)

  O *Open* quer dizer que as informações sobre a existência, a natureza e a localização do conteúdo são abertas e podem ser conhecidas

- *Archives* (arquivos) – significa o armazenamento de conteúdos em computadores – os conteúdos estão "arquivados" nos computadores

Open Archives Initiative Protocol for Metadata Harvesting

"The Open Archives Initiative Protocol for Metadata Harvesting (referred to as the OAI-PMH) provides **an application-independent interoperability framework based on metadata harvesting**."

"O Protocolo de Coleta de Dados da Iniciativa de Arquivos Abertos (referenciado como OAI-PMH) oferece um **ambiente de interoperabilidade, baseado na coleta de metadados, que é independente das aplicações** (computadores, seus programas e suas tecnologias)."

OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting:

- P = *protocol* = protocolo = maneira convencionada de executar uma tarefa
- M = *metadata* = metadados
- H = *harvesting* = colheita ou coleta

OAI-PMH – Open Archives Initiative Protocol for Metadata Harvesting:

É uma maneira convencionada de recolher, de forma automatizada(*), metadados de arquivos abertos (que abrem seus metadados mas não necessariamente os seus conteúdos)

(*) Se não fosse automatizada, seria inviável frente ao número de repositórios que existem e a quantidade de conteúdos que abrigam!!

Para que serve o OAI-PMH?

O OAI-PMH serve para transferir metadados de um sistema para outro, independentemente das tecnologias e produtos utilizados para implementar os sistemas

O **arquivo aberto** é o que **disponibiliza seus metadados** para serem coletados e disponibilizados em outros repositórios.

Os conteúdos podem ser protegidos e sujeitos a condições de uso – seja por pagamento ou por pertencer a uma comunidade ou, ainda, por um sigilo temporário (p.ex. aguardando a publicação em um periódico ou o registro de uma patente).

Do que é composto o OAI-PMH?

- Metadados – a informação a ser coletada
- Formatos dos registros – a estrutura da informação a ser coletada (em xml)
- Repositórios – os locais onde a informação a ser coletada estará disponível

- Identificadores – expressões que permitem identificar de forma única cada registro no repositório – os identificadores são um dos problemas mais sérios nas bibliotecas digitais
- Política de coleta – definição lógica para estabelecer o que será coletado, por exemplo, por data ou por tipo de informação

- Instruções (*verbs*) que as máquinas entendem e que permitem a extração da informação (são 6 verbos) – *GetRecord*, *Identify*, *ListIdentifier*, *ListMetadataFormats*, *ListRecords*, *ListSets*

Para que o protocolo funcione, são necessários dois atores:

- Provedores de dados – aqueles sistemas que possibilitam a busca a seus dados através do Protocolo, ou seja, são os que "abrem" seus arquivos
- Provedores de serviço – aqueles sistemas que recolhem os dados dos outros e constroem bases integradas com maiores facilidades de acesso ao público (agregando valor ao processo)

- Algumas instituições podem ser ambos, ou seja, funcionam como provedores de dados em algumas situações e como provedores de serviços em outras

BDTD – Biblioteca Digital de Teses e Dissertações:

- Catálogo coletivo de metadados de teses e dissertações brasileiras *online* em texto completo
- Coordenada e operada pelo IBICT
- Possui metadados de 12 instituições, num total superior a 11 mil registros

- O IBICT coleta os metadados das bibliotecas digitais das universidades – é o provedor de serviços
- As universidades abrem seus arquivos para que o IBICT colete os metadados – são as provedoras de dados
- O IBICT abre os arquivos da BDTD para que a NDLTD e o CyberTesis Net coletem os metadados das teses e dissertações e os levem para outros catálogos coletivos de metadados – nesta situação o IBICT é provedor de dados

Passos para implementar o protocolo:

- Preparação – definições:
  - Metadados a serem coletados
  - Formato dos registros (xml *schema*)
  - Política de coleta
  - Freqüência de coleta

- Preparação – desenvolvimento dos programas:
  - Provedores de dados – para extraírem os metadados de suas bases no formato convencionado e entregarem os provedores de serviços durante a coleta
  - Provedores de serviços – para acionarem os provedores de dados, receberem os metadados e levarem aos catálogos coletivos

- Operação:
  - O provedor de serviços aciona os provedores de dados com a periodicidade e a política de coleta estabelecidas
  - O provedor de dados fornece os metadados ao provedor de serviços
  - Os metadados são levados ao catálogo coletivo
  - Todas as operações são realizadas de forma automatizada, sem intervenção humana

O catálogo coletivo (base cooperativa) é centralizado e independe:

- Da tecnologia dos sistemas provedores de dados
- Dos metadados (adicionais) usados localmente em cada provedor de dados
- De haver controle sobre o uso dos conteúdos
- Dos conteúdos serem gratuitos ou pagos

O catálogo coletivo:

- Tem a finalidade de localização dos conteúdos
- Não se propõe à descrição completa dos conteúdos
- Não armazena e não realiza transações com os conteúdos
- Passa todas as transações às bibliotecas digitais dos provedores de dados

Finalidade geral:

Possibilitar encontrar (busca e recuperação) um conteúdo e transferir o acesso a cada situação local

Referências

[01] BDTD – Biblioteca Digital de Teses e Dissertações
http://bdtd.ibict.br/bdtd/

[02] CyberTesis Net
http://www.cybertesis.net/

[03] Lagoze, Carl & Sompel, Herbert Van de
The Open Archives Initiative: building a low-barrier interoperability framework
JCDL 2001 – Joint Conference on Digital Libraries
http://www.openarchives.org/documents/jcdl2001-oai.pdf

[04] NDLTD - Networked Digital Library of Theses and Dissertations
http://www.ndltd.org/

[05] Open Archives Initiative
http://www.openarchives.org/