

# Uso do software livre Solr para construção da camada de indexação e busca do módulo de consulta institucional do sistema Ainfo

Arthur Eziquiel Moraes do Nascimento<sup>1</sup>  
Isaque Vacari<sup>2</sup>

## Introdução

O Ainfo é um sistema para gestão de bibliotecas e recuperação de informações que permite a gestão da informação técnico-científica, integrando bases de dados documentais (obras impressas ou registradas em outros suportes físicos, bem como de publicações eletrônicas), cadastrais e de processos bibliográficos. O seu desenvolvimento e evolução são de responsabilidade da Embrapa Informática Agropecuária com o acompanhamento da Comissão Permanente para o Ainfo (CP-Ainfo) (EMBRAPA INFORMÁTICA AGROPECUARIA, 2010).

Desde sua primeira versão, em 1991, o software vem sendo evoluído com os objetivos de implementar novas funcionalidades requeridas por seus usuários e de atualizá-lo tecnologicamente, em especial no que se refere a novas ferramentas de software. É utilizado amplamente pelas bibliotecas da Embrapa. A versão atual em uso nas bibliotecas é a 6.2.2 (Ainfo 6), cuja plataforma tecnológica consiste de acesso via web ao sistema utilizando-se do navegador de Internet Firefox.

---

<sup>1</sup> Universidade Estadual de Campinas; [arthur@cnptia.embrapa.br](mailto:arthur@cnptia.embrapa.br)

<sup>2</sup> Embrapa Informática Agropecuária; [isaque@cnptia.embrapa.br](mailto:isaque@cnptia.embrapa.br)

Para a nova versão do sistema Ainfo concebeu-se a construção de um módulo específico de consulta institucional, com acesso restrito à Embrapa, para uso dos bibliotecários e usuários da biblioteca. A concepção do novo módulo de consulta, levou o projeto a um estudo de tecnologias livres para recuperação de informação com recursos mais sofisticados de indexação e busca em bases de dados.

## Objetivos

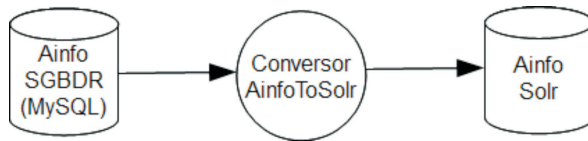
O mecanismo de indexação e busca textual Lucene (THE APACHE SOFTWARE FOUNDATION, 2010a) tem sido largamente utilizado em projetos *open source* em todo mundo. Entretanto, para o módulo de consulta do sistema Ainfo, objetivou-se encontrar soluções livres capazes de estender e melhorar os recursos originais da ferramenta Lucene, como: integração com bases de dados relacionais, recursos mais sofisticados de indexação e busca, e maior integração com a arquitetura Java EE (*Java Platform, Enterprise Edition*).

## Material e métodos

O resultado dessa investigação levou a escolha do software livre Solr. (THE APACHE SOFTWARE FOUNDATION, 2010b). O Solr é um mecanismo de indexação e busca textual de código aberto, mantido pela *Apache Software Foundation*, baseado na biblioteca Lucene que oferece recursos sofisticados de indexação e busca textual, como: busca com operadores booleanos, busca específica por campo, *highlighting* sobre o resultado da busca, paginação do resultado da busca, *facets* sobre o resultado da busca (recurso presente em sítios web de comércio eletrônico), *caching* de busca, integração com banco de dados relacionais, replicação de bases de dados, interface de administração web etc. (SMILEY; PUGH, 2009).

O Solr é disponibilizado em uma aplicação web, o sistema cliente pode ser construído com as linguagens de programação: Ruby, PHP, Java, Python, .NET, C# e Perl.

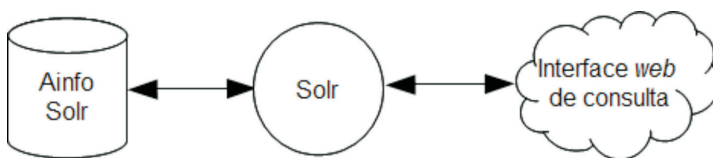
A primeira tarefa de uso do Solr no sistema Ainfo foi converter os dados das tabelas: Acervo Documental (com aproximadamente 850.000 registros), Periódicos (com aproximadamente 41.300 registros) e Bibliotecas (42 registros) para o formato de busca suportado pelo Solr. Para realizar a tarefa mencionada, foi construído um software de conversão específico denominado AinfoToSolr, cuja arquitetura está relatada na Figura 1.



**Figura 1.** Arquitetura de conversão de dados do formato Ainfo - MySQL para o Ainfo - Solr.

O software AinfoToSolr é responsável por converter os dados do sistema Ainfo armazenados no Sistema Gerenciador de Banco de Dados Relacional MySQL para o formato de dados entendido pela ferramenta Solr.

Após a conclusão do software AinfoToSolr, o projeto tem avançado no desenvolvimento da interface web de busca. Essa nova interface permitirá uma consulta sofisticada e integrada a todo acervo documental e digital das 42 bibliotecas da Embrapa. A Figura 2 apresenta a arquitetura do módulo de consulta do sistema Ainfo.



**Figura 2.** Arquitetura do módulo de consulta do sistema Ainfo.

## Resultado e discussão

O uso da ferramenta Solr mostrou-se adequado para a construção da camada de indexação e busca do módulo de consulta institucional do sistema Ainfo, bem como para conquista de autonomia tecnológica em software livre para recuperação de informação. Por fim, o uso de software livre tem-se consolidado nas atividades de pesquisa, desenvolvimento e inovação da Embrapa.

## Referências

EMBRAPA INFORMÁTICA AGROPECUÁRIA. Ainfo. 2010. Disponível em: <<http://www.ainfo.cnptia.embrapa.br>>. Acesso em: 12 ago. 2010.

SMILEY, D.; PUGH, E. Solr 1.4 Enterprise Search Server. Birmingham, Mumbai: Packt Publishing, 2009. 336 p.

THE APACHE SOFTWARE FOUNDATION. Apache Lucene. 2010a. Disponível em: <<http://lucene.apache.org>>. Acesso em: 12 ago. 2010.

THE APACHE SOFTWARE FOUNDATION. Apache Solr. 2010b. Disponível em: <<http://lucene.apache.org/solr>>. Acesso em: 12 ago. 2010.