

TÉCNICAS DE RECOMENDAÇÃO E RECUPERAÇÃO DE CONTEÚDO PARA REPOSITÓRIOS INSTITUCIONAIS BASEADO NA UNIVERSIDADE FEDERAL DE LAVRAS (UFLA)

Thiago Bellotti Furtado
Universidade Federal de Lavras
Caixa Postal 3037 – Lavras, MG, Brasil
55 35 3829-1178
thiagobellotti@gmail.com

Ahmed Ali Abdalla Esmín
Universidade Federal de Lavras
Caixa Postal 3037 – Lavras, MG, Brasil
55 35 3829-1597
aesmin@gmail.com

Nivaldo Oliveira
Universidade Federal de Lavras
Caixa Postal 3037 – Lavras, MG, Brasil
55 35 3829-1174
nivaldo@biblioteca.ufla.br

INTRODUÇÃO

Iniciativas relacionadas a democratização da produção científica tem alcançado uma grande importância na sociedade de pesquisa. Uma dificuldade encontrada nesse processo está relacionada a disponibilização dessas informações. Por isso, tem-se discutido sobre qual melhor meio a ser adotado para disseminar, armazenar e preservar as informações a fim de permitir o acesso permanente e confiável da produção científica. Para suprir essa demanda, a utilização de repositórios institucionais tem sido amplamente adotada nas universidades em âmbito nacional e internacional. Com objetivo de compartilhar e disseminar o conhecimento produzido, os repositórios auxiliam na promoção do acesso à informação por um processo mais transparente e democrático, aumentando o impacto do reconhecimento e da visibilidade nacional e internacional das produções científicas desenvolvidas nas instituições.

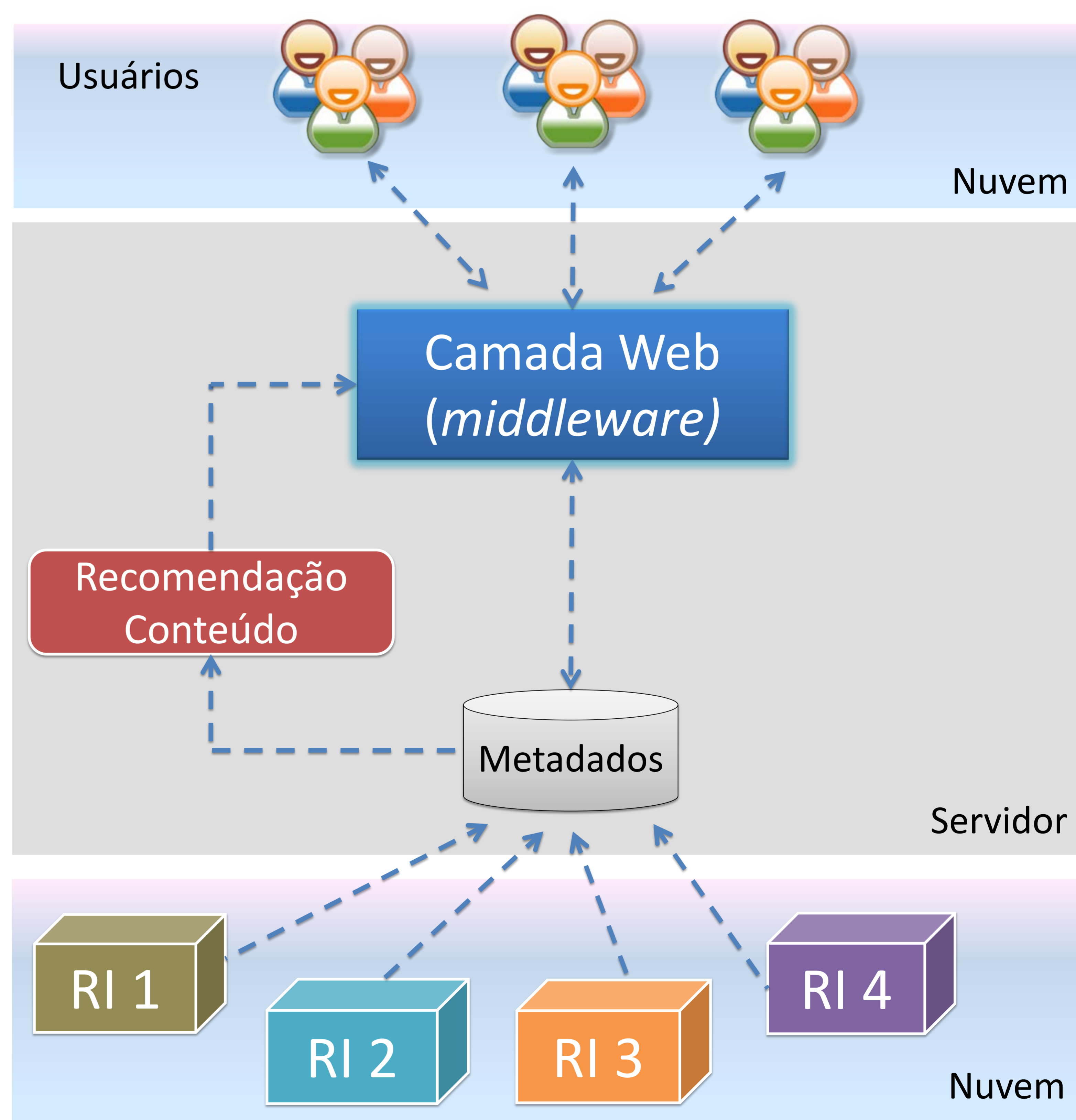
A Biblioteca Universitária da Universidade Federal de Lavras – UFLA através da parceria com o Instituto Brasileiro de Informação em Ciência e Tecnologia - IBICT e a Financiadora de Estudos e Projetos - FINEP implantou o Repositório Institucional da UFLA (RIUFLA), que foi construído através do software DSpace, atualmente utilizado para manutenção e gerenciamento de seu repositório. O DSpace é um software sem fins lucrativos utilizado para construção de repositórios digitais abertos. Além de ser um sistema *open source*, sua estrutura é completamente personalizável permitindo ser adaptado de acordo com a necessidade da instituição. O software permite fácil acesso a maioria dos tipos de conteúdo digital, incluindo texto, imagens e vídeos. Sua arquitetura é formada por três camadas: camada de aplicação, de armazenamento e de negócios. A camada de armazenamento é implementada usando um sistema de arquivos que é gerenciado através de um banco de dados. Na camada de negócios são definidas as funcionalidades, incluindo fluxo de trabalho, gerenciamento, administração e pesquisa de conteúdo e a camada de aplicação abrange a interface de apresentação web do sistema.

OBJETIVOS

Com intuito de ampliar a capacidade de pesquisa dos repositórios institucionais, procura-se criar uma nova camada (*middleware*) que permita a interoperabilidade entre os repositórios de forma transparente. Além disso, para melhorar a exploração do conteúdo nos repositórios, é proposto um estudo sobre técnicas de mineração de dados direcionadas a recomendação de conteúdo sobre o Repositório Institucional da UFLA.

METODOLOGIA

A coleta de dados dos repositórios poderá ser feita pelo módulo *REST API (Jersey)* utilizado para recuperação de metadados. As informações recuperadas serão estruturadas formando uma base de dados gerenciada pelo software PostgreSQL, mantido em servidor Web Apache contendo também o DSpace. A camada (*middleware*) de comunicação (interface Web) consulta a base de metadados retornando resultados das pesquisas dos usuários. Técnicas de recomendação de conteúdo podem ser aplicadas sobre essas informações para refinar a busca e sugerir conteúdo de interesse do usuário.



RESULTADOS ESPERADOS

O propósito deste trabalho é implementar uma camada (*middleware*) para possibilitar a comunicação e recuperação de informações entre os repositórios desenvolvidos sobre o DSpace. Com essa camada será possível recuperar os metadados de diversos repositórios, centralizando a busca de informações para o usuário em uma única aplicação. Para facilitar e filtrar as informações recuperadas, técnicas de recomendação de conteúdo serão aplicadas sobre a base de dados gerada, a fim de sugerir documentos e permitir que as informações selecionadas sejam direcionadas ao interesse específico dos usuários que utilizam os repositórios.