

Uma proposta para a filtragem de informação em bibliotecas digitais

Diego F. Ucha Centro Universitário Senac e FAPESP (diegoucha@gmail.com)

Marcelo Furusawa Honório Centro Universitário Senac

Fabício Jailson Barth Centro Universitário Senac

Orlando Rodrigues Júnior Centro Universitário Senac

CONTEXTO E MOTIVAÇÃO

Garantir o acesso à informação tem se tornado um assunto em destaque na sociedade. Ter acesso à informação certa no momento adequado é pré-requisito para muitas situações do dia-a-dia. Por outro lado, a quantidade de informação disponível – especialmente na Internet – cresce exponencialmente, dificultando a recuperação eficiente da informação. Não diferentemente, as bibliotecas digitais também passam pelos mesmos problemas no que se refere ao acesso à informação.

Uma vez que as abordagens convencionais de disponibilização da informação ainda não são eficazes, outras linhas de ação devem ser exploradas. Uma alternativa promissora é utilizar técnicas da área da inteligência artificial a partir de perfis de usuários que serão utilizados na recomendação personalizada de documentos.

PROPOSTA

O serviço proposto por este projeto irá classificar um documento como relevante ou não para o usuário de uma biblioteca digital. As recomendações geradas serão disponibilizadas aos usuários através do envio de e-mail, por meio de feeds RSS e em áreas específicas da biblioteca digital com recomendações para documentos distintos. Este serviço é composto por duas partes: um mecanismo para criação e manutenção do perfil do usuário e um mecanismo para recomendação de documentos.

CRIAÇÃO E MANUTENÇÃO DO PERFIL DO USUÁRIO

A construção do perfil do usuário será feita de maneira implícita para o usuário. O perfil será criado com base em suas preferências individuais de documentos relevantes, identificadas através do 'Mapeamento de ações', e em qual(is) grupo(s) de usuários ele se encaixa, identificado(s) através da correspondência de documentos relevantes entre os usuários.

MAPEAMENTO DE AÇÕES

1. Acesso direto ao documento
2. Acesso a documento similar, referenciado ou citado
3. Imprimir, salvar em PDF ou exportar em XML
4. Enviar por e-mail
5. Adicionar à coleção do usuário
6. Como referenciar este documento
7. Exportar referência ao documento em formato específico
8. Ver estatísticas de acesso
9. Monitoramento de citação ou estatísticas de acesso

RECOMENDAÇÃO DE DOCUMENTOS

Para realizar a recomendação será utilizada uma abordagem híbrida que faz uso de técnicas de recomendação baseadas nos conteúdos acessados por um usuário e pelo grupo de usuários que ele pertence (técnica colaborativa). Depois de aplicadas as duas técnicas no mesmo conjunto de dados, os resultados de ambas são unificados. Ao utilizar uma abordagem híbrida pode-se minimizar os problemas encontrados em cada uma das abordagens quando utilizadas de forma isolada.



TÉCNICA BASEADA EM CONTEÚDO

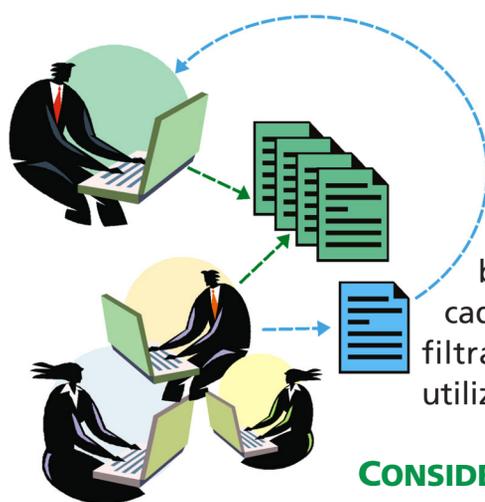
A fórmula utilizada para descrever um documento é a TF-IDF e a utilizada para obter a afinidade de um perfil de um usuário com a descrição de um documento é a do Cosseno.

FÓRMULA TF-IDF

$$wt_i(d) = \left| tf_i \cdot \log \left| \frac{N}{nt_i} + 0.5 \right| \right| / \sqrt{\sum_{j=1}^n (tf_j)^2 \cdot \log^2 \left| \frac{N}{nt_j} + 0.5 \right|}$$

FÓRMULA DO COSSENO

$$Sim(u, d) = \cos(u, d) = \frac{\sum_{i=1}^n wt_i(u) \cdot wt_i(d)}{\sqrt{\sum_{j=1}^n wt_j(u) \cdot \sum_{j=1}^n wt_j(d)}}$$



TÉCNICA COLABORATIVA

O algoritmo para identificar a similaridade de documentos, com base dos documentos relevantes a cada um dos usuários é o algoritmo de filtragem colaborativa item por item, utilizado pela Amazon.com.

CONSIDERAÇÕES FINAIS

A solução apresentada está alinhada com a Web 3.0, que utiliza técnicas de inteligência artificial empregadas para melhorar a qualidade dos serviços disponíveis on-line.

Através da personalização do serviço, o sistema proposto possibilita melhorar o acesso à informação em bibliotecas digitais de uma forma eficiente.