

Sistema Automático de Disseminação Seletiva de Informação

Marcello Peixoto Bax¹, Lídia Alvarenga², Fernando Silva Parreiras³,
Wladimir Cardoso Brandão⁴

[bax, lidiaalvarenga, fparreiras, web] @eci.ufmg.br

Resumo

A quantidade de usuários de bases de dados bibliográficas está crescendo rapidamente. Com esta alta taxa de crescimento e dada a dificuldade de manter-se atualizados, torna-se essencial para os usuários destas bases ter acesso aos últimos documentos incluídos. É desejável que os usuários enviem seus perfis, isto é, consultas que são avaliadas constantemente, para que possam ser automaticamente informados de novas inclusões de seu interesse. Este serviço é tradicionalmente chamado de disseminação seletiva da informação (DSI). O crescimento da base de artigos e da base de usuários são exemplos de desafios enfrentados por um DSI. Ao invés de ser a parte ativa no acesso de informação, os usuários desejam ser notificados sempre que um novo artigo em uma dada área da pesquisa for publicado. Neste artigo, propõe-se um modelo para DSI em um projeto de biblioteca digital em desenvolvimento na escola da ciência de informação da UFMG. Descrevem-se os objetivos e principais serviços que esta infra-estrutura fornecerá.

Palavras-chave: SABiO, disseminação seletiva, bibliotecas digitais, agentes inteligentes

Abstract

The number, size, and user population of bibliographic and full-text document databases are rapidly growing. With a high document arrival rate, it becomes essential for users of such databases to have access to the very latest documents; yet the high document arrival rate also makes it difficult for users to keep themselves updated. It is desirable to allow users to submit profiles, i.e., queries that are constantly evaluated, so that they will be automatically informed of new additions that may be of interest. Such service is traditionally called Selective Dissemination of Information (SDI). The high document arrival rate, the huge number of users, and the timeliness requirement of the service pose a challenge in achieving efficient SDI. Instead of being the active part in information access, researchers want to be notified whenever a new paper in one's research area is published. In this article, we propose a model for SDI in a digital library project in development at the School of Information Science at UFMG. We describe the goals and main services this infrastructure will provide.

¹ Professor do Programa de pós-graduação em ciência da informação da ECI / UFMG

² Professora do Programa de pós-graduação em ciência da informação da ECI / UFMG

³ Mestrando em ciência da informação pela ECI / UFMG

⁴ Mestrando em ciência da informação pela ECI / UFMG

Keywords: SABiO, selective dissemination, digital libraries, intelligent agents.

1. Introdução

Os avanços das tecnologias de informação e comunicação (TICs) têm contribuído para o acelerado aumento da quantidade de informações disponíveis hoje. A maior disponibilidade de recursos de informação demanda serviços de intermediação que permitam a coleta e filtragem automáticas das fontes e a distribuição seletiva desse conteúdo aos usuários de acordo com seu perfil de interesse. Trata-se aqui do campo de pesquisas dos Sistemas de Recuperação de Informação (SRI).

Calvin Mooers (Mooers 1951) define três problemas a serem equacionados nesse campo: o da representação e organização da informação, o da especificação da busca por informação e o da criação de mecanismos para recuperação. Saracevic mostra que existem diversas abordagens desenvolvidas para a modelagem do usuário (Saracevic, Spink, Wu 1997), atividade relacionada ao problema de especificação da busca por informação. As abordagens estão centradas no sistema, como é o caso da “*relevance feedback*” (Spink, Losee 96) e da expansão de consultas “*query expansion*” (Efthimiadis 1996), ou no usuário, como a “*question analysis*” (Taylor, 1968), ou no sistema e no usuário, como os sistemas de disseminação seletiva (Luhn 1961).

O processo de Disseminação Seletiva de Informações (DSI), na língua inglesa “*selective dissemination of information (SDI)*”, foi concebido por Hans Peter Luhn, da IBM Corporation, em 1958, com a finalidade de aperfeiçoar serviços de alerta oferecidos por bibliotecas, centros de documentação e sistemas especializados de informações documentais. A grande expansão da literatura mundial, a proliferação das bases referenciais especializadas on-line e a necessidade de atualização dos usuários das inúmeras especialidades da comunidade científica e tecnológica foram fatores motivadores desse tipo de serviço automatizado de informação já nos anos 60.

Luhn define DSI como “...*serviço que consiste em direcionar novos itens de informação, de qualquer que seja a fonte, para aqueles pontos onde a probabilidade de seu uso, em conexão com o interesse corrente [do usuário], seja alta*”. (Luhn apud Housman, 1973). Como se vê, o tema não é recente. Porém, sua rediscussão, à luz das possibilidades inovadoras oferecidas pelas bibliotecas digitais e a tecnologia de redes de comunicação Web, se justifica cada vez mais, diante da quantidade de informação disponível e da necessidade crescente em se manter atualizado.

O objetivo desse artigo é propor um modelo de arquitetura funcional a ser integrado a um projeto de concepção de infra-estrutura de apoio à construção de bibliotecas digitais, em desenvolvimento na Escola de Ciência da Informação da UFMG. Sua contribuição é especificar em alto nível o funcionamento de um módulo de DSI para futura implementação do serviço no projeto SABiO (Sistemas de Agentes para Bibliotecas Digitais). Além de discutir o DSI no ambiente atual de redes Web e suas novas possibilidades.

Na Seção 2 apresentam-se os conceitos que cercam o tema DSI, além de um breve histórico das principais iniciativas realizadas no Brasil. Os trabalhos relacionados a partir

de revisão atualizada da literatura internacional são apresentados na Seção 3. A Seção 4 descreve o interesse da adição de um módulo de DSI na Arquitetura SÁBIO. O modelo a ser adotado é especificado na Seção 5 pelo levantamento dos requisitos de arquitetura, funcionalidade e desempenho desejáveis. Finalmente, na Seção 6, são discutidas algumas questões à guisa de conclusão.

2. Disseminação Seletiva de Informação

Housman (1973), em capítulo publicado no ARIST, explora as maneiras pelas quais o conceito de DSI vinha sendo implementado nos anos que antecedem a publicação de seu trabalho. Na essência, segundo o autor, o DSI seria a automação de uma função clássica da biblioteca tradicional, que consiste em informar ao usuário sobre as novas aquisições, compatíveis com seu interesse de consulta.

Um típico serviço de DSI tem por objetivo prover cada usuário, inscrito com uma lista periódica e personalizada, dos novos trabalhos que deram entrada na base de dados e que podem se constituir em subsídios para trabalhos em andamento ou interesses. Assim, cada usuário inscrito recebe um diferente conjunto de informações referenciais, dependendo de seus interesses particulares, tal como definido previamente, em seu respectivo “perfil de interesse”.

Um perfil pode ser visto como uma consulta que permanece no sistema e continua a produzir resultados com o passar do tempo e a medida em que novos recursos são adicionados ao sistema. No processo de seleção das informações e no levantamento dos perfis, os sistemas podem utilizar vocabulários livres ou controlados (tesauros e/ou ontologias). As estratégias de buscas, assim como nos atuais mecanismos de busca na Web, são formuladas por meio de operadores lógicos (E, OU e NÃO), permitindo a utilização de peso, truncagem e limitação de perfis por idioma, dependendo do modelo de DSI implementado.

Em alguns países os serviços de DSI se iniciaram na década de 60, expandindo-se nas décadas posteriores. Tornaram-se serviços correntes de disseminação nos grandes sistemas de recuperação de informações referenciais, das instâncias públicas (governamentais) e privadas (empresariais e comerciais). No Brasil a literatura registra algumas iniciativas relevantes, dentre as quais destacam-se:

a) O SONAR, Sistema Orientado para Notificação Automática de Referências, criado em 1972 pela Comissão Nacional de Energia Nuclear, Centro de Informações Nucleares, CIN/CNEN, Rio de Janeiro, utilizando a bases de dados INIS, *International Nuclear Information System*, vinculado e coordenado à Agência Internacional de Energia Atômica – IAEA, versão impressa denominada ATOMINDEX, cobrindo as áreas de ciência físicas, química, geociências, ciências biológicas, engenharia e tecnologia, economia, legislação e documentação. O CIN/CNEN gerou uma tecnologia de recuperação utilizando um sistema de pesos que atua na seleção de referências a partir de termos extraídos de um mini-tesauro apoiado no tesauro do INIS;

b) No final da década de 70 o Departamento de Informação e Documentação (DID) da Empresa Brasileira de Pesquisa Agropecuária (EMBRAPA) iniciou um serviço automatizado de DSI, após uma experiência prévia de disseminação manual. Inicialmente o DID operou apenas com a base de dados AGRICOLA da *National Agricultural Library* dos Estados Unidos, incorporando ao serviço posteriormente as seguintes bases de dados: *Commonwealth Agricultural Bureaux (CAB)*; *International Food Information System (IFIS)*; BIOSIS, composto por *Biological Abstracts* e *Bio-Research Index*; *Chemical Abstracts Service*. Criado para atender os pesquisadores da empresa, esse serviço foi mais tarde estendido aos pesquisadores de outras instituições de ensino e pesquisa do setor agrícola brasileiro, chegando a atender também a entidades de países do Cone Sul;

c) O Instituto de Pesquisas Tecnológicas (IPT) implantou um sistema automatizado de DSI destinado a disseminar informações constantes da base de dados COMPENDEX, produzida pela *Engineering Index Inc.* dos Estados Unidos.

O fascículo número 2, volume 6 do periódico Revista de Biblioteconomia de Brasília foi dedicado à temática DSI, com artigos relatando experiências na Faculdade de Medicina Veterinária e Zootecnia da USP, na Companhia Vale do Rio Doce, no Centro de Informação Científica para a Saúde da Secretaria de Estado de Saúde do Rio de Janeiro, disseminação seletiva de informações para os alunos de pós-graduação em Saúde Pública e Administração Hospitalar, experiência em biblioteca universitária e detalhamentos relativos às experiências do CIN/CNEN e da EMBRAPA.

3. Trabalhos Relacionados

Os sistemas de DSI podem ser classificados como baseados em conteúdo, em colaboração ou híbridos. Nesses sistemas o perfil pode ser tanto informado explicitamente pelo usuário ao sistema quanto construído a partir da experiência de utilização do usuário (suas buscas, anotações e comentários associados aos recursos).

Os sistemas **baseados em conteúdo** (ou de filtragem por conteúdo), determinam a relevância de um recurso e realizam o casamento (*matching*⁵) entre o perfil do usuário e o recurso pela análise de seu conteúdo.

Os sistemas **baseados em colaboração** (filtragem colaborativa) determinam a relevância de um recurso realizando o *matching* entre o perfil do usuário e perfis de usuários com interesses similares ou perfis comunitários padrão (Ferreira, Silva 2001). Vários autores, tais como Peter Wurman, Shardanand e Pattie Maes além de Michael Wellman, analisam sistemas de recomendação seguindo a abordagem “filtragem colaborativa” em ambientes comerciais, como Amazon.com e outros.

⁵ Ato de verificar repetição ou igualdade entre conjuntos de dados.

Dentre os sistemas DSI baseados em conteúdo destacam-se as iniciativas de Tak Yan, denominada SIFT (*Stanford Information Filtering Tool*) (Yan, Garcia-Molina 1999) e de Ken Lang, denominada NewsWeeder (Lang 1995). O SIFT foi um dos primeiros serviços DSI na Internet. Começou de maneira experimental em fevereiro de 1994 e se tornou um serviço comercial em abril de 1996 lidando, naquela época, com mais de 18.400 usuários, 40.100 perfis e processando cerca de 80.000 documentos por dia. Suas fontes de informação são notícias da USENET e artigos de diversas listas de correio eletrônico. O NewsWeeder, apesar de possuir alguns aspectos de DSIs colaborativos, é considerado um DSI baseado em conteúdo. Foi submetido a testes de desempenho em 1994 e se mostrou superior a outras iniciativas de DSI.

Quanto à forma de captação, filtragem e armazenamento de recursos, os serviços funcionam similarmente: periodicamente varrem fontes de informação, recuperam as referências para recursos novos ou alterados, realizam a indexação de cada um, armazenam seus descritores em repositório de dados e, finalmente, realizam o *matching* dos descritores com os perfis de usuários, disseminando, a partir dos resultados da comparação, os recursos ou suas referências de acesso.

A diferença está na forma de obtenção e gerenciamento dos perfis de usuário. Enquanto no SIFT os perfis são construídos explicitamente pelo próprio usuário, através de correio eletrônico ou de formulários Web, no NewsWeeder o perfil é construído por navegação - definido pelo sistema através das experiências de navegação passadas. A idéia do modelo de navegação do NewsWeeder pressupõe que, ao navegar pelos recursos em busca de informações, o usuário indique seu nível de interesse pelo recurso em uma escala de pontos de 1 a 5, fazendo com que o sistema aprenda suas preferências e defina seu perfil de consulta automaticamente.

É importante ressaltar que, embora ambos os sistemas expressem os perfis utilizando o modelo vetorial, a atribuição de pesos ocorre de maneira diferente. O SIFT segue os princípios *tf-idf* (*term-frequency/inverse-document-frequency*) que observam que: 1) quanto mais vezes um termo aparece em um documento, mais relevante ele é no contexto do documento e; 2) quanto mais vezes um termo aparece no conjunto de todos os documentos, menos ele é capaz de discriminar os documentos desse conjunto. No SIFT, os termos com baixo poder de discriminação não são eliminados, mas sim têm seu grau de importância minimizado. Já o NewsWeeder segue o princípio MDL (*Minimum Description Length*), procurando eliminar do modelo termos com baixo poder de discriminação. De acordo com Lang (Lang 1995), o método MDL apresentou melhor desempenho se comparado com as técnicas que utilizam *tf-idf*.

No universo dos serviços de DSI baseados em colaboração têm-se o GroupLens (Resnik, et al. 1994). Integrável à arquitetura do sistema USENET, o GroupLens é um sistema distribuído para disseminação da informação que utiliza regras e cálculos de determinação de níveis de interesse de grupos de usuários em documentos de uma coleção para prever os interesse de outros usuários e grupos. Em uma arquitetura distribuída, os servidores colhem os níveis de interesse dos usuários, através de um modelo de navegação similar ao utilizado pelo NewsWeeder. Além disso, compartilham

esses níveis de interesses com outros servidores utilizando técnicas de propagação (herdadas da USENET) e troca de mensagens. Elaboram também tabelas de predição de interesses, através da combinação dos diferentes níveis de interesses indicados pelos usuários, que são utilizadas para definir o grau de relevância dos documentos para eles. Na elaboração das tabelas de predição, as opiniões de cada usuário assumem pesos diferentes ditados pela sua volatilidade histórica de opinião: as opiniões de usuários que historicamente mantêm seus critérios de julgamento de valor possuem peso maior sobre as opiniões daqueles mais voláteis. Sua força está na capacidade de predição da relevância de um documento a um determinado usuário com base na identificação de interesses comuns entre ele e outros grupos que também fazem uso do sistema.

Independentemente da categoria, é crescente o número de iniciativas de construção de serviços de DSI. E como em todo campo em crescimento, é necessário o estabelecimento de bases conceituais sólidas que garantam a avaliação das iniciativas e permitam avanços contínuos. Nesse sentido, os pesquisadores portugueses João Ferreira e Alberto Silva propuseram em 2001 o MySDI (Ferreira, Silva 2001). Com o objetivo de oferecer uma arquitetura genérica para suportar o projeto e a construção de serviços de DSI, o modelo conceitual MySDI fornece balizes para a construção de sistemas em conformidade com os requisitos de arquitetura, funcionalidade e desempenho desejáveis nesses tipos de serviços.

Podem-se citar muitos outros sistemas de DSI presentes na literatura, tais como o HERMES (Faensen, 2001) e o SIENA (Carzaniga, 2000), além de outros.

4. Modelo Conceitual SABiO-DSI

A concepção central por trás da integração de agentes de DSI à arquitetura SABiO é a de que os usuários de bibliotecas digitais desejam manter-se atualizados de forma mais conveniente lançando mão de um serviço de alerta, além de buscar e navegar em suas coleções favoritas. Com efeito, trabalhos no âmbito do projeto HERMES revelam que serviços de alerta que integram diferentes provedores de informação são indispensáveis para os usuários (Faensen et al., 2001), já que os alivia da tediosa tarefa de buscar e navegar ou até mesmo inscrevendo-se a vários sistemas individuais de alerta.

Nesse contexto, o objetivo dos agentes SABiO-DSI é sugerir recursos de informação que interessam aos usuários. Para tal, eles precisam inferir o gosto e o interesse dos usuários, explorando diversas fontes de informação explícitas e implícitas.

4.1. Integração na arquitetura SABiO

SABiO é uma arquitetura para construção de bibliotecas digitais baseada na metáfora dos sistemas multi-agentes. Tais arquiteturas são uma alternativa atualmente popular aos sistemas mais centralizados do tipo cliente-servidor. Em sua forma mais pura, sistemas multi-agentes não possuem servidores ou funcionalidades centrais. Todos os nós da rede são considerados componentes hierarquicamente iguais. Alguns dos benefícios das

arquiteturas multi-agentes são: adaptação, balanceamento de carga (escalabilidade), tolerância à falhas e capacidade de integrar diversas fontes de informação distribuídas, entre outras.

Nos últimos anos, sistemas multi-agentes com arquitetura *peer-to-peer* (P2P) se tornaram populares em aplicações de compartilhamento de arquivos (cf. Projeto Edutella). A arquitetura SABiO, tal como relatada por Bax (1997), é uma arquitetura multi-agente, porém não P2P em seu modelo mais puro, totalmente descentralizado. A arquitetura do SABiO pode ser considerada como híbrida, entre Cliente/Servidor e P2P.

Arquiteturas baseadas em Agentes, como o SABiO, oferecem as características desejáveis para se implementar sistemas multi-agentes com suporte a consultas, perfis e notificações. Atualmente existem pelo menos duas opções que podem ser utilizadas para implementar o serviço de notificação especificado para integrar a arquitetura SABiO. Uma utilizaria a biblioteca de códigos (API) da plataforma de agentes desenvolvida no projeto DIET, e a outra utilizaria a API do projeto *JXTA* da *Sun Microsystems* (www.jxta.org). DIET é um projeto para conceber uma plataforma multi-agente de código livre aberta, robusta, adaptativa e escalonável (Marrow et. al., 2001; Hoile et. al., 2002; DIET Agents Website).

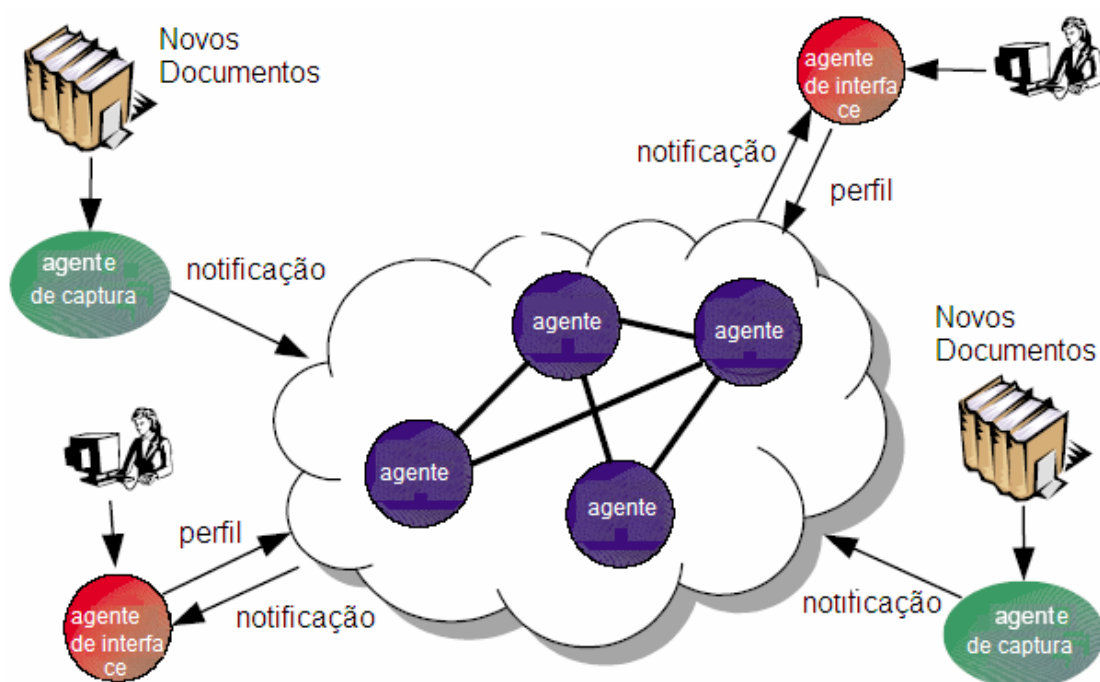


Figure 1 - Arquitetura multi-agente do SABiO-DSI.

Com a integração do módulo de agentes DSI à arquitetura SABiO espera-se preencher uma lacuna hoje existente nas arquiteturas multi-agente aplicadas a construção de bibliotecas digitais. Os sistemas P2P descritos na literatura ou possuem capacidades de consulta pelo usuário, i.e., Gnutella (Gnutella e Edutella), ou apenas tratam

perfis/notificações, i.e., SIENA (Carzaniga, 1998). A especificação do serviço DSI na arquitetura SABiO procura unificar essas abordagens (consulta e notificação) em um ambiente de construção de bibliotecas digitais. Vale notar ainda que tais sistemas (Gnutella, Edutella e SIENA) não foram desenvolvidos originalmente para serem utilizados com foco em bibliotecas digitais.

4.2. O SABiO-DSI

SABiO-DSI é um modelo conceitual cujo objetivo é servir de base para se comparar e discutir sistemas de DSI que seguem configurações de arquiteturas multi-agente. Outro objetivo é servir de referência para a implementação de um serviço de DSI, integrando à arquitetura de agentes da plataforma SABiO.

O diagrama de contexto da Figura 2 ilustra a integração à plataforma SABiO, mostrando os atores que interagem com o modelo SABiO-DSI, ou seja, os usuários e os outros agentes SABiO.

Como mostra a Seção 3, os principais problemas a serem solucionados em um sistema DSI são: 1) a especificação das necessidades de informação do usuário – seu perfil; e 2) o casamento de padrão (*matching*) eficiente entre os novos recursos que entram na base e o perfil dos usuários – a filtragem. As próximas seções descrevem, em detalhes, como esses dois problemas são tratados no modelo.

4.2.1. Agentes de interface SABiO-DSI

Os agentes de interface gerenciam informações que caracterizam as necessidades de informação do usuário no longo prazo (manutenção dos perfis), além de realizar o envio de notificações. Os perfis são utilizados como base para a busca e notificação automáticas. Os perfis alimentam o serviço de notificação automática, que envia alertas aos usuários de acordo com o surgimento de novos recursos de seu interesse.

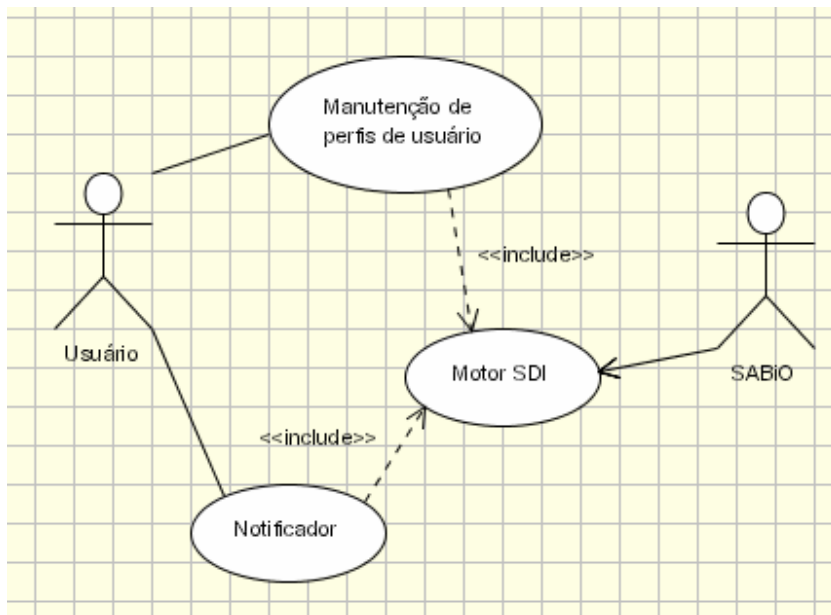


Figure 2 - Diagrama de contexto do SABiO-DSI.

O SABiO-DSI utiliza dois mecanismos de interação com o usuário: 1) Interface Web: onde o usuário tem uma postura ativa e interage com o sistema navegando em suas telas, criando, enviando e armazenando consultas e informações sobre seu perfil; 2) Envio de mensagens por correio eletrônico: o usuário recebe uma notificação sobre eventos relacionados ao seu perfil. Estes eventos podem ser novos artigos (os usuários são informados da inclusão de novos artigos, baseado na associação da classificação do documento com a classificação eleita pelo usuário) ou novos usuários (comunicação de incorporação de novos usuários que possuam um perfil semelhante).

O perfil do usuário é construído diretamente pelo próprio usuário através do preenchimento de formulários. A partir destas respostas, o sistema faz a associação dos dados fornecidos com os dados da base, colocando o usuário em um segmento já existente ou criando um novo, caso ainda não exista perfis semelhantes.

O SABiO-DSI armazena as seguintes informações sobre os perfis dos usuários:

- a) Frequência de notificação: periodicidade em que o usuário recebe informações do sistema.
- b) Atributos do perfil do usuário: nome, escolaridade, área de interesse, residência, etc.
- c) Palavras-chave eleitas pelo usuário: para descrever seus interesses.
- d) Consultas gravadas pelo usuário: consultas avançadas que utilizam operadores lógicos.

4.2.2. Agentes de Associação e Filtragem (Motor DSI):

Esta camada é responsável por identificar similaridades entre perfis dos usuários utilizando os atributos armazenados e realizar, a partir destes, a construção de comunidades ou grupos de interesses comuns. Além disso, responde pelo armazenamento

de consultas. O usuário pode optar por armazenar ou não a consulta, e pela sua utilização ou não para notificação, transformando-a em uma consulta permanente. Dessa forma, as consultas são consideradas atributos do perfil dos usuários. Todos os descritores utilizados pelo usuário em suas consultas são gravados pelo sistema. Os usuários podem consultar quais são os descritores mais utilizados, dando-lhes a opção de considerá-los ou não para fins de notificação.

A implementação dos agentes de Associação e Filtragem será baseada nos trabalhos de Yan e Garcia-Molina, que, em 1994, definiram a estrutura de índices para indexação de perfis e algoritmos eficientes para o casamento de padrões entre documentos e perfis (Yan e Garcia-Molina, 1994). Estas estruturas originaram o sistema SIFT já citado.

Como visto acima na Seção 4.2.1, os usuários expressam suas necessidades de informação em perfis, gerenciando-o pela interface Web. Um perfil é uma consulta e uma regra de notificação. Enquanto a consulta especifica o conteúdo que o usuário deseja receber, a regra de notificação especifica como o conteúdo será entregue, ou seja, com que frequência (diariamente, semanalmente), através de qual protocolo (e-mail) e formato (HTML, texto puro, BibTeX).

As consultas não são diferentes daquelas que podem ser submetidas na interface de busca do SABiO, que são classificadas em simples ou avançadas. As consultas avançadas podem usar operadores booleanos como expressão de filtragem. Uma consulta avançada é uma expressão booleana formada de pares atributo/valor e separados por operadores lógicos. e.g., “author=Marcos de Andrade” or “title LIKE '%disseminação%' ”. O conjunto de atributos são os campos bibliográficos (metadados) extraídos quando um documento dá entrada em uma coleção do SABiO. Um valor de ranking expresso em percentagem.

Dados bibliográficos que casam com a consulta em níveis superiores àqueles expressos no registro de ranking da consulta são registrados para serem enviados ao usuário. Alternativamente, os usuários podem especificar que os N registros bibliográficos mais relevantes devem ser enviados. O agente de filtragem submete a consulta conforme a periodicidade escolhida pelo usuário e escolhe os documentos que superam a adequação ao ranking e possuem data de entrada no sistema posterior à data do último envio.

4.2.3. Camada de Agentes do SABiO

Esta camada não será descrita neste artigo, e é representada na Figura 2 como um ator na representação UML. É citada apenas para melhor contextualizar o SABiO-DSI. Além de outros agentes, ela consiste nos agentes responsáveis por coletar os recursos para a base SABiO.

5. Conclusão

Um requisito importante para serviços de DSI é a personalização, que por sua vez, depende de interação privilegiada com o usuário, que deve prover informações explicitamente ou implicitamente pela própria utilização do sistema. Nesse último caso, o sistema observa o comportamento de uso do usuário.

Ambientes de bibliotecas digitais se beneficiam de formas de interação até então inexistentes nos serviços de bibliotecas tradicionais, passando a ser interessante que tais ambientes ofereçam serviços de DSI.

Dessa forma, espera-se com esse artigo iniciar uma discussão de idéias que não são recentes, mas que, quando aplicadas em novos ambientes de bibliotecas digitais podem representar fonte de inovação, com resultados favoráveis para o desenvolvimento da ciência e tecnologia nacionais. Assim, propomos aqui um modelo conceitual para a discussão do tema e também com base de especificação de requisitos para a concepção de um DSI integrado à arquitetura SABiO.

A motivação para se aplicar a DSI no SABiO é a criação de funcionalidades que vão além da busca e navegação e propõem a notificação dos usuários baseada em filtragem de novos documentos da base por perfil.

Nesse contexto, esse trabalho tem pela frente os seguintes desafios técnicos:

- a) Estudo de uma variedade de modelos e linguagens de consultas para DSI textual que utilizam técnicas de banco de dados, recuperação de informação e lógica formal.
- b) Desenvolvimento de algoritmos de filtragem de informação usando indexação dos perfis de usuários.
- c) Implementações de protocolos para interação entre os agentes da plataforma SABiO.
- d) Testes reveladores de robustez, escalabilidade e eficiência desta plataforma.

6. Referencial Bibliográfico

Bax, M. P. (1997) *Agentes de Interface para Bibliotecas Digitais: a arquitetura SABiO*, VI SEAB, Águas de Lindóia, SP.

Carzaniga, A. (1998) *Architectures for an Event Notification Service Scalable to a Wide Area Networks*. Politecnico di Milano, PhD thesis, Italy.

Carzaniga, A.; Rosenblum D. S.; Wolf A. L. (2000) *Achieving scalability and expressiveness in an Internet-scale event notification service*. Proceedings of the nineteenth annual ACM symposium on Principles of distributed computing. Disponível em: < <http://doi.acm.org/10.1145/343477.343622> >. Acesso em: 30 mai. 2004.

DIET Agents website. <http://diet-agents.sourceforge.net>.

Eduella Project. <http://edutella.jxta.org>.

Efthimiadis, E. N. (1996) *Query expansion*. In WILLIAMS, M. Annual Review in Information Science and Technology, v.31, p.121-187.

Faensen, D. et al. (2001) *Hermes – A Notification Service for Digital Libraries*. In: JC'DL'01, Roanoke, Virginia, USA.

Ferreira, J.; Silva, A. (2001) *MySDI: A Generic Architecture to Develop SDI Personalised Services (How to Deliver the Right Information to the Right User?)*. In: Proceedings of the ICEIS'2001. Setubal, Portugal. p.262-270. Disponível em: < <http://citeseer.ist.psu.edu/489601.html> >. Acesso em: 25 mai. 2004.

Gnutella website. <http://www.gnutella.com>.

Hoile, C. et. al. (2002) *Core specification and experiments in DIET: a decentralised ecosystem-inspired mobile agent system*. Proceedings of the 1st International Joint Conference on Autonomous Agents & Multiagent Systems (AAMAS 2002), Bologna, Italy. Disponível em: < <http://citeseer.ist.psu.edu/565145.html> >. Acesso em: 30 mai. 2004.

Housman, Edward M.(1973). *Selective dissemination of information*. Annual Review of Information Science and Technology. Washington, American Society of Information Science, p. 221-241.

Lang, K. (1995) *NewsWeeder: learning to filter netnews*. In: Kaufmann, M. Proceedings of the 12th International Conference on Machine Learning. Lake Tahoe, CA. Disponível em: < <http://citeseer.ist.psu.edu/lang95newsweeder.html> >. Acesso em: 25 mai. 2004.

Luhn, H. P. (1961) *Selective dissemination of new scientific information with the aid of electronic processing equipment*. American Documentation, v.12, p.131-138.

Marrow P. et. al. (2001) *Agents in Decentralised Information Ecosystems: The DIET Approach*. Proceedings of the AISB'01 Symposium on Information Agents for Electronic Commerce, AISB'01 Convention, University of York, United Kingdom. Disponível em: < <http://www.intelligence.tuc.gr/publications/aisb01.pdf> >. Acesso em: 30 mai. 2004.

Mooers, C. (1951) *Zatocoding applied to mechanical organization of knowledge*. American Documentation, v.2, p.20-32.

Revista de Biblioteconomia de Brasília. (1978) Brasília, Associação dos Bibliotecários do Distrito Federal, ABDF; Departamento de Biblioteconomia da Faculdade de Estudos Sociais Aplicados da Universidade de Brasília, v.6, n.2.

Saracevic, T.; Spink, A.; Wu, M. (1997) *User and intermediaries in information retrieval: what are they talking about?* In: Jameson, A.; Paris, C.; Tasso, C. User Modeling: Proceedings of the Sixth International Conference, UM97. Vienna, New York: Springer Wien New York. Disponível em: <<http://wwwis.win.tue.nl:8080/2ID10/Resources/users-and-intermediaries.pdf>>. Acesso em: 25 mai. 2004.

Spink, A.; Losee, R. M. (1996) *Feedback in information retrieval*. In Williams, M. Annual Review in Information Science and Technology, v.31, p.33-78.

Taylor, R. S. (1968) *Question negotiation and information seeking in libraries*. College & Research Libraries, v.29, p.178-194.

Yan, T. W.; Garcia-Molina, H. (1994) *Index structures for selective dissemination of information under the Boolean model*. Transactions on Database Systems (TODS), v.19, n.2. Disponível em: <<http://doi.acm.org/10.1145/176567.176573>>. Acesso em: 25 mai. 2004.

Yan, T. W.; Garcia-Molina, H. (1999) *The SIFT information dissemination system*. ACM Transactions on Database Systems (TODS), v.24, n.4, p.529-565. Disponível em: <<http://citeseer.ist.psu.edu/yan00sift.html>>. Acesso em: 25 mai. 2004.