UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA

2010.1

PREFERÊNCIAS DO USUÁRIO PARA PERSONALIZAÇÃO DE CONSULTAS

TRABALHO DE GRADUAÇÃO

Aluno: Gustavo Ferraz Diniz (gfd@cin.ufpe.br)
Orientadora: Ana Carolina Salgado (acs@cin.ufpe.br)

Recife, 22 de julho de 2010

UNIVERSIDADE FEDERAL DE PERNAMBUCO

GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO
CENTRO DE INFORMÁTICA

2010.1

GUSTAVO FERRAZ DINIZ

PREFERÊNCIAS DO USUÁRIO PARA PERSONALIZAÇÃO DE CONSULTAS

Monografia apresentada ao Centro de Informática da Universidade Federal de Pernambuco como requisito parcial para obtenção do Grau de Bacharel em Ciência da Computação.

Orientadora: Ana Carolina Salgado

Recife, 22 de julho de 2010

Agradecimentos

Agradeço primeiramente a Deus, por ter me dado forças nos momentos difíceis.

À minha família, em especial meus pais, por me apoiarem e me incentivarem em todas as etapas da minha vida.

À minha orientadora, a professora Ana Carolina Salgado, pela oportunidade e confiança para trabalhar neste projeto.

Aos professores do CIn, em especial ao professor Fernando Fonseca, pela experiência e conhecimento passados nos dois anos de atividade de monitoria.

Ao meu parceiro neste projeto, o aluno de mestrado Thiago Arruda, pelo apoio durante o desenvolvimento deste trabalho.

Aos meus amigos pelo companheirismo e apoio prestado durante o curso.

Agradeço, enfim, a todos que de alguma forma contribuíram para a realização deste trabalho.

Resumo

A personalização de consultas consiste no processo de dinamicamente melhorar uma consulta de acordo com preferências relatadas pelos usuários, produzindo respostas distintas para cada perfil. Este trabalho tem como objetivo aperfeiçoar o módulo de submissão e reformulação de consultas do sistema SPEED (*Semantic Peer-to-Peer Data Management System*), utilizando o conceito de personalização. Será desenvolvido um mecanismo que permita o armazenamento e a captura do perfil do usuário. Com base nos perfis informados as consultas submetidas deverão ser personalizadas de maneira que uma mesma consulta retorne diferentes resultados para cada usuário.

Palavras-chaves: Preferências; Personalização; Consulta; SPEED.

Abstract

Query Personalization is the process of dynamically enhancing a query according to user preferences stored in a user profile in order to provide personalized answers. This work aims to improve the SPEED's (*Semantic Peer-to-Peer Data Management System*) query module by using the concept of personalization. It intends to develop a mechanism for capturing and storing the user profile. The search results are personalized based on profiles in a way that users may find different relevant things to search due to their different preferences.

Keywords: Preferences; Personalization; Query; SPEED.

Sumário

Lista de Figuras

1. Introdução

Neste capítulo será feita uma introdução do contexto no qual foi desenvolvido o trabalho. Serão abordados aspectos como motivação ao desenvolvimento do projeto e seus objetivos, além da estrutura desta monografia.

1.1 Motivação

A popularização da Internet tem ocasionado um grande aumento na produção e disseminação da informação. Entretanto, o grande volume de dados e novidades está tornando o exercício da pesquisa cada vez mais importante e difícil. A tarefa de encontrar informações relevantes e úteis é trabalhosa, mesmo sabendo onde procurar e por onde começar [18].

Um usuário que acessa um sistema de informação com a intenção de satisfazer uma necessidade de conhecimento sobre um determinado assunto, poderá ter que reformular a consulta várias vezes [12]. Freqüentemente, é necessário incluir termos na pesquisa pertinentes ao domínio, pensando em diversos sinônimos para ampliar as chances de encontrar conteúdos relevantes e filtrar os resultados obtidos de forma manual, extraindo todos os itens que não traduzem o seu real interesse [14]. Esta é uma experiência muito comum, especialmente para usuários da web, devido à abundância de informações e a heterogeneidade dos conteúdos existentes.

A fim de prover soluções que permitam, da melhor maneira possível, filtrar o que é mais relevante e/ou de interesse do usuário, arquiteturas e técnicas de personalização vêm sendo abordadas em pesquisas recentes. Esta área de estudo tem sido denominada de personalização de consultas, e consiste no processo de dinamicamente melhorar a consulta de acordo com preferências relatadas pelos usuários, produzindo respostas personalizadas [10].

O ponto chave desse estudo é que usuários distintos poderão obter resultados diferentes mesmo que efetuem a mesma consulta, uma vez que será considerada a heterogeneidade de habilidade, conhecimento e comportamento dos usuários.

1.2 Objetivos

Dentro do contexto apresentado na seção anterior, esse trabalho tem como objetivo aperfeiçoar o módulo de submissão e reformulação de consultas do sistema SPEED (*Semantic Peer-to-Peer Data Management System*) [24], utilizando o conceito de personalização. Será desenvolvido um mecanismo que permita o armazenamento e a captura do perfil do usuário, contendo informações que foram explicitamente definidas pelo próprio. Com base nos perfis informados as consultas submetidas deverão ser personalizadas de maneira que uma mesma consulta retorne diferentes resultados para cada usuário. Além daquelas características informadas pelo usuário através de seu perfil, o sistema também poderá levar em consideração informações do contexto do usuário, como localização, por exemplo, refletindo na classificação ou exclusão de resultados para uma dada consulta que for submetida.

1.2 Estrutura da Monografia

Além deste capítulo introdutório, que explanou sobre motivação e objetivos, esta monografia encontra-se organizada em mais quatro capítulos.

O Capítulo 2 define os conceitos relativos às preferências do usuário, personalização de consultas e modelagem de usuários, listando algumas características importantes de cada conceito. Em seguida, é discutido o que está sendo pesquisado na área de preferências e personalização, através da análise de trabalhos correlatos.

O Capítulo 3 faz uma introdução ao SPEED, apresentando alguns de seus fundamentos, e mostra como este trabalho contribuiu para que o módulo de submissão de consultas desse sistema fosse estendido com um mecanismo de armazenamento e captura do perfil do usuário, objetivando a personalização das consultas.

O Capítulo 4 aborda os detalhes da implementação da estrutura de personalização das consultas no SPEED, destacando ferramentas utilizadas durante o desenvolvimento e explorando as telas de interface com usuário.

Por fim, o Capítulo 5 descreve as conclusões adquiridas com o trabalho aqui apresentado, além de definir sugestões para trabalhos futuros.

2. Preferências do Usuário

Neste capítulo serão apresentados alguns conceitos e características relacionados às preferências do usuário. Além disso, será discutido como as preferências podem estar relacionadas com os conceitos de personalização de consultas e modelagem de usuários. Por fim, será descrito o que está sendo pesquisado nessa área, através da análise de trabalhos correlatos.

2.1 Definição

De acordo com o dicionário Aurélio da língua portuguesa [5], o termo preferência pode ser definido como o ato de preferir uma pessoa ou uma coisa a outra. Trata-se de um conceito bastante utilizado nas ciências sociais, particularmente na Economia [11]. Ele assume uma escolha real ou imaginária entre alternativas e a possibilidade de ordenar essas alternativas, baseado na felicidade, satisfação, gratificação, prazer ou utilidade que fornece. Em ciências cognitivas, preferências individuais permitem a escolha de objetivos [27].

As preferências do usuário podem ser estudadas segundo duas abordagens [25]: a abordagem numérica ou quantitativa e a abordagem simbólica ou qualitativa. Na primeira, funções de utilidade calculam o grau (ranking) das alternativas e o de maior valor calculado é selecionado como solução. Já o enfoque qualitativo ou simbólico permite a explicitação de uma ordem parcial sobre as preferências, o que facilita a compreensão e a explicação das mesmas [21]. As Figuras 1 e 2 são exemplos de preferências na abordagem quantitativa e qualitativa, respectivamente.

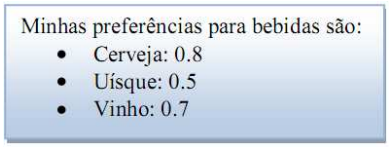
Minhas preferências para bebidas são:
- Cerveja: 0.8
- Uísque: 0.5
- Vinho: 0.7

Figura 1 – Exemplo de preferência quantitativa [21]

A desvantagem do enfoque quantitativo é que muitas vezes além de não ser necessário expressar preferências sobre todos os elementos, alguns deles podem ser naturalmente incomparáveis. O enfoque qualitativo se assemelha mais como realmente as pessoas expressam preferências, uma vez que não existem funções de utilidade ou probabilidades associadas às alternativas.

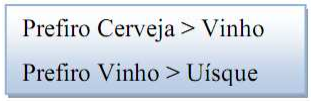

Figura 2 – Exemplo de preferência qualitativa [21]

O tratamento das preferências do usuário está se tornando cada vez mais importante na composição dos sistemas de informação atuais [2]. Em um ambiente onde existe sobrecarga de informação, verifica-se que a busca de assuntos relevantes aos interesses dos usuários está se tornando uma tarefa morosa e complexa [14]. Dessa forma, as preferências podem ser utilizadas para personalizar as buscas e realizar a filtragem da informação, reduzindo o volume de dados que devem ser apresentados para o usuário.

2.2 Personalização de Consultas

Como mencionado anteriormente, em sistemas onde existe um grande volume de informações disponíveis, a realização de consultas objetivas pode se transformar em uma tarefa bastante difícil. Nesse contexto, as preferências do usuário representam o conceito chave que permitirá a construção de consultas personalizadas. A partir de uma personalização eficiente, os resultados das pesquisas realizadas poderiam ser ofertados de acordo com as necessidades dos usuários, reduzindo, portanto, os esforços empregados para a localização de conteúdo relevante.

Segundo Koutrika [12], personalização de consulta consiste no processo de dinamicamente melhorar a consulta de acordo com preferências relatadas pelos usuários, através de seu perfil, com o objetivo de produzir respostas mais objetivas e, conseqüentemente, respostas mais enxutas. A idéia consiste em que pessoas com opiniões e comportamentos distintos obtenham diferentes resultados para uma mesma pesquisa.

O conceito de personalização de consultas foi abordado por diversos trabalhos com diferentes objetivos [10, 11, 12, 13, 25]. Entretanto, é possível identificar nos estudos três categorias básicas para personalização: baseada em perfil, baseada em comportamento e baseada em colaboração.

A personalização baseada em perfil faz o uso das preferências e características pessoais dos usuários. Em uma abordagem simplificada, as características são obtidas

implicitamente ou através da declaração explicita em formulários ou questionários realizados pelo próprio usuário.

A personalização baseada em comportamento permite que as consultas sejam adequadas a partir de resultados solicitados anteriormente. Esta técnica faz uso de registros que contêm as ações realizadas pelos usuários em cada sessão de utilização do serviço. O Google, por exemplo, disponibiliza uma ferramenta [7] de consulta que leva em conta o histórico de navegação do usuário, podendo ser considerada uma personalização baseada em comportamento.

A personalização baseada em colaboração, bastante comum em sistemas de recomendação, faz uso da navegação social, que permite que os usuários deixem rastros de navegação, como opiniões ou votação, para que outros usuários possam utilizar durante as suas consultas [13]. O motor de busca Eurekster [4] é um exemplo da aplicação da personalização colaborativa.

Quando se trata o assunto personalização, por envolver informações de terceiros, os seguintes pontos devem ser considerados [1]: privacidade e ética, segurança e confiança na informação apresentada. Isso se deve ao fato de se tratar de informações pessoais dos usuários. Esses dados devem ser trabalhados de modo seguro e sempre com o conhecimento e consentimento de quem os fornece.

2.3 Modelagem de Usuários

A modelagem de usuários refere-se ao processo de construção de um modelo, no qual são representadas explicitamente as características e preferências de um usuário ou de um grupo de usuários [14].

De acordo com Santos [20], a construção de modelos de usuário é motivada pela possibilidade dos sistemas personalizarem o seu comportamento conforme as necessidades de cada usuário. O modelo pode ser constituído por informações como nome, idade, país, nível educacional, profissão, interesses, informações geográficas e outras características relevantes ao sistema em questão [13].

Segundo Micarelli [16], os modelos podem ser considerados dinâmicos, quando o perfil é atualizado constantemente, conforme a interação do usuário com o sistema, ou estáticos, quando as informações se mantêm inalteradas ao longo do tempo.

A tarefa de obtenção do perfil do usuário pode acontecer de forma explícita, implícita ou híbrida [13]. A coleta de características de forma explícita requer a atuação direta do usuário para indicar as suas preferências. Freqüentemente, tais informações são capturadas através do preenchimento de formulários, onde os usuários expressam opiniões que podem ser utilizadas para diversos fins, como definir o grau de interesse por determinado assunto. Esta técnica requer a pré-disposição do usuário, uma vez que depende do tempo e vontade do mesmo para disponibilizar os seus dados.

A coleta de preferências de forma implícita é realizada através do monitoramento das atividades dos usuários por parte do sistema. A abordagem implícita requer um menor esforço do usuário, sendo, normalmente, necessária a utilização de um agente de software que acompanha a interação do mesmo com o sistema para identificar os seus interesses.

Tanto a abordagem implícita quanto explícita apresenta vantagens e desvantagens [13]. A primeira permite construir e atualizar o perfil por meio da interação com o sistema, porém não consegue identificar claramente características negativas. A segunda consegue capturar opiniões mais detalhadas, entretanto depende da disponibilidade do usuário de oferecer as informações corretamente. Diante disso, a utilização das duas técnicas em conjunto, na abordagem híbrida, pode facilitar a construção de um perfil mais amplo para cada usuário.

Ainda de acordo com Micarelli [16], em sistemas de personalização de consultas, a modelagem do usuário pode ser aproveitada de três formas distintas: personalização como parte do processo de recuperação, re-classificação de resultados e reformulação de consultas.

Na primeira forma, mostrada pela Figura 3 (a), a classificação é um processo unificado, na qual o perfil do usuário é utilizado para pontuar os resultados encontrados. Nesta técnica, é provável que os resultados para consulta sejam retornados mais rapidamente, uma vez que o mecanismo tradicional de classificação pode ser diretamente adaptado para incluir a personalização, evitando desperdício de processo computacional.

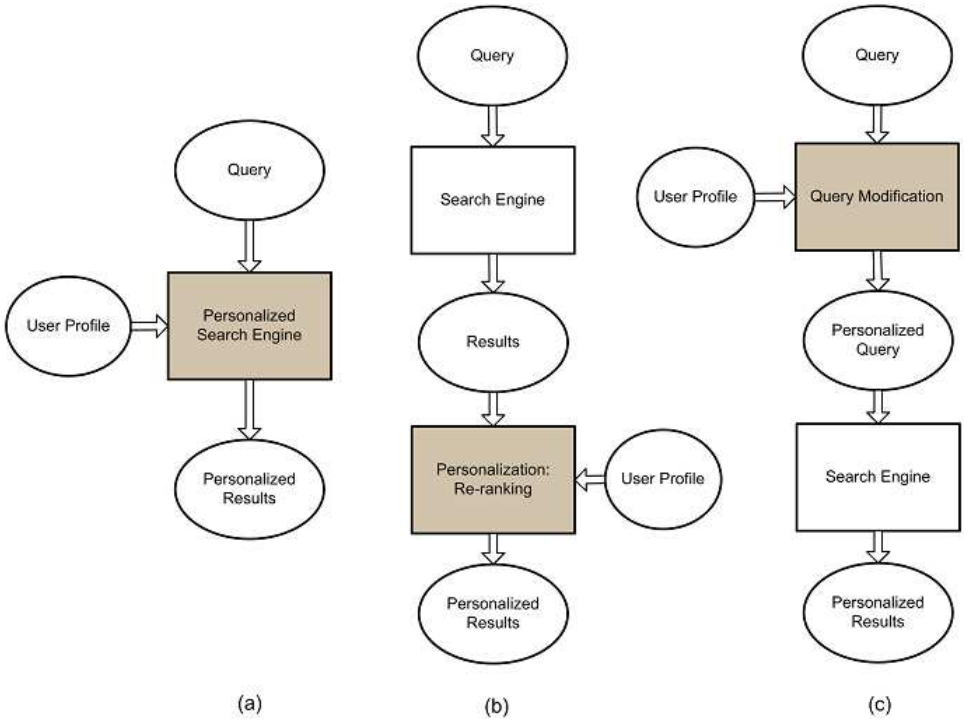

Figura 3 – Processo de personalização como parte do processo de recuperação (a), por re-classificação dos resultados (b) ou por reformulação da consulta (c) [16]

A segunda forma, apresentada na Figura 3 (b), re-classifica os resultados da consulta a partir do perfil do usuário. Esta técnica funciona em uma segunda etapa no processo, uma vez que inicialmente a consulta tradicional é realizada e, a partir dos resultados retornados, são classificados os objetos através da pontuação personalizada. Muitos sistemas implementam essa abordagem no lado cliente da aplicação, onde o software se conecta a um serviço de busca e os resultados encontrados são analisados localmente.

A terceira e última abordagem realiza uma reformulação da consulta propriamente dita. Esta técnica utiliza o perfil do usuário para modificar a *string* de busca original, realizando um enriquecimento da consulta. Portanto, a consulta poderá ser transformada através da adição ou remoção de palavras-chaves que melhor representem as necessidades do usuário. A Figura 3 (c) esquematiza esta terceira forma.

2.4 Trabalhos Relacionados

Conforme citado anteriormente, na literatura, os conceitos de preferências e personalização foram introduzidos por diversos trabalhos. Tais estudos estão distribuídos pelas mais diferentes áreas, tratando objetivos distintos. Dentre os temas mais abordados nas pesquisas de personalização de consultas, pode-se destacar a reformulação de consultas em banco de dados e a personalização para buscas na Web.

Koutrika e Ioannidis [10] desenvolveram um framework de personalização para sistemas de bancos de dados baseado nos perfis dos seus usuários. Eles observaram que uma consulta submetida a um banco de dados retorna sempre as mesmas informações, independentemente das preferências da pessoa que realizou a consulta. Assim, foi definida uma modelagem para os usuários do banco, utilizando tanto a abordagem explícita quanto implícita. A partir dos perfis gerados, foi possível modificar a estrutura das consultas, adicionando algumas restrições que representavam as necessidades do usuário.

A arquitetura geral apresentada pelo trabalho, mostrada na Figura 4, inclui vários módulos em torno do módulo tradicional responsável pelo acesso aos dados. O sistema mantém um repositório de informações dos usuários (*User Profiles*), que pode ser alimentado explicitamente pelo usuário ou através do monitoramento das interações do mesmo com o sistema (*Profile Creation*). As características dos perfis são utilizadas tanto no módulo de reformulação de consultas (*Query Personalization*), quanto no módulo de apresentação do resultado (*Presentation Personalization*).

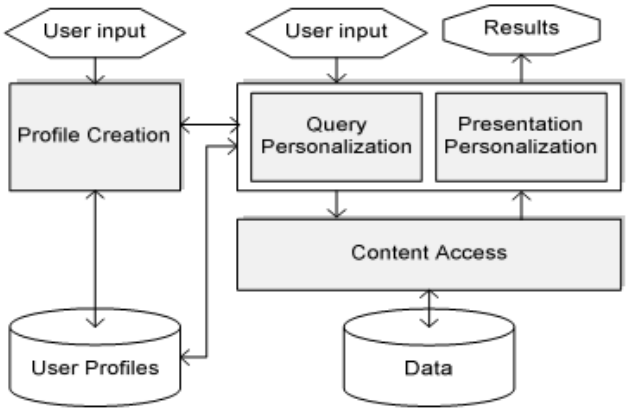

Figura 4 – Arquitetura para personalização de consultas em banco de dados [10]

Ainda seguindo a linha de personalização de consultas em banco de dados, o próprio Koutrika apresentou em outro estudo [11], uma proposta visando aperfeiçoar o modelo discutido anteriormente. Nesse outro trabalho, ele acrescenta o conceito de colaboração para personalização das consultas. A questão levantada é que as pessoas freqüentemente fazem escolhas não apenas baseadas em suas preferências pessoais, mas também nas características e opiniões de outras pessoas.

Outra temática bastante comum nos trabalhos sobre preferência e personalização é a questão referente a buscas personalizadas na Web. Esses trabalhos procuram cada vez mais melhorar a experiência e satisfação do usuário em suas consultas.

Marchi [14] chama a atenção para o fato de que a sobrecarga de informações presentes na web está tornando a busca de conteúdo relevante aos interesses dos internautas uma atividade dispendiosa e complexa.

Micarelli [16] levanta a questão que apesar dos motores de busca convencionais possuírem capacidade de retornar resultados de boa qualidade em resposta à maioria das consultas, eles ainda não conseguem oferecer os resultados de maneira eficiente. O autor defende que a utilização de modelos de usuários baseados no perfil, histórico de consultas e colaboração, traz resultados mais satisfatórios que os engenhos com técnicas tradicionais de recuperação de informação.

Outro trabalho de Marchi [13] propõe uma arquitetura para um sistema de personalização de busca na Web que emprega a técnica de indexação de semântica latente, adaptada para o ambiente Web, em conjunto com um modelo de usuário construído de forma implícita por meio do acompanhamento da navegação do mesmo nos documentos resultantes da busca. A técnica de semântica latente permite identificar a relação semântica entre as páginas, refletindo na ordenação dos resultados. Já a modelagem do usuário oferece condições de identificar os interesses do usuário na busca, possibilitando uma melhoria na ordenação dos resultados de acordo com esses interesses. A arquitetura proposta nesse trabalho é apresentada na Figura 5.

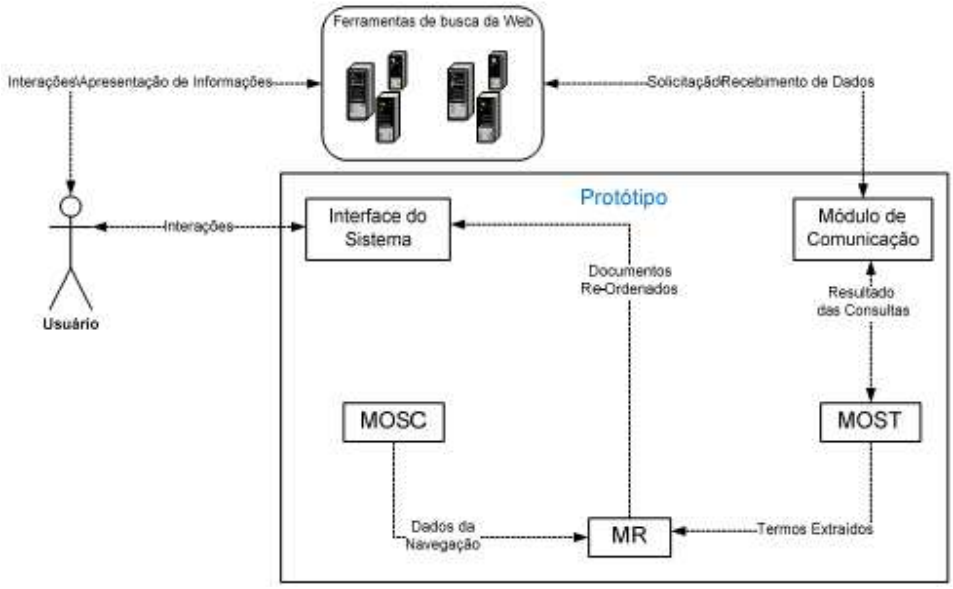

Figura 5 – Arquitetura para personalização de consultas Web [13]

A arquitetura sugere a composição de cinco módulos inter-relacionados: interface do sistema, comunicação, operações sobre as consultas (*MOSC*), operações sobre o texto (*MOST*) e re-classificação (*MR*).

O módulo de interface tem o objetivo de fornecer o apoio às ações dos usuários, apresentando os resultados das re-ordenações de acordo com os processamentos realizados pelo sistema proposto.

O módulo de comunicação estabelece a ligação entre a aplicação e os sistemas externos, solicitando as informações necessárias e retornando os resultados obtidos ao módulo de re-classificação.

O módulo de operações sobre texto se encarrega de capturar e processar os termos dos documentos identificados pelo *MOSC*, armazenando os termos em uma estrutura de dados e realizando os processamentos necessários para a extração do conteúdo.

O módulo de operações sobre as consultas é responsável por identificar os interesses do usuário nos resultados obtidos pelas pesquisas. Essa identificação é capturada por meio da navegação, sendo consideradas as ações do mouse e o tempo de permanência em cada documento navegado.

Por fim, o módulo de re-classificação é o responsável por re-ordenar o conteúdo de acordo com a similaridade entre o modelo de usuário construído pelo *MOSC* e as páginas retornadas.

Já no trabalho de Giannopoulos [6], é apresentado um framework para melhorar os processos de ordenação e re-ordenação de páginas na personalização de buscas na web. O método discutido é baseado na quantidade de cliques dos usuários nas páginas, com o objetivo de criar múltiplas funções de ordenação correspondentes a diferentes assuntos de pesquisa. Essas funções são combinadas, cada vez que um usuário realiza uma nova busca, para produzirem uma nova ordenação, levando em consideração a similaridade da consulta com assuntos pesquisados anteriormente.

2.5 Considerações

Neste capítulo foram discutidos conceitos relacionados às preferências do usuário, personalização de consultas e modelagem de usuários. O entendimento desses conceitos é muito útil para a compreensão do contexto no qual o projeto foi desenvolvido. Também foram abordados alguns trabalhos que aliam preferência à personalização, como forma de enriquecer a experiência do usuário.

No próximo capítulo, será abordada a utilização das preferências do usuário para a personalização de consulta em um ambiente dinâmico e distribuído como o SPEED.

3. Personalização de Consultas no SPEED

Este capítulo descreve como o sistema SPEED pode associar o conceito de preferências do usuário com personalização de consultas. Inicialmente, serão discutidas algumas características básicas relacionados ao SPEED como, por exemplo, a arquitetura atual do seu módulo de consulta. Em seguida, serão propostas novas funcionalidades que definem um mecanismo de armazenamento e captura das preferências dos usuários com o objetivo de permitir a personalização de consultas.

3.1 Definições de PDMS e SPEED

Um PDMS (*Peer Data Management Systems)* é um sistema de gerenciamento de dados em ambientes *peer-to-peer* (P2P) que oferece uma infra-estrutura adequada para manipular informações armazenadas em fontes distribuídas, heterogêneas e autônomas [19]. Estes sistemas possibilitam o compartilhamento de dados estruturados e/ou semi-estruturados entre os pontos (*peers*), oferecendo representações semânticas mais ricas para os dados compartilhados.

Do ponto de vista de gerenciamento de dados, um PDMS deve tratar questões como localização dos dados e processamento de consulta, além da integração e consistência das informações [26].

O sistema SPEED (*Semantic Peer-to-Peer Data Management System*) [24] é um PDMS baseado em semântica que tem como objetivo prover facilidades para o compartilhamento e a integração dos dados distribuídos ao longo dos pontos conectados.

Assim como vários tipos de PDMS, o principal serviço oferecido pelo SPEED é o processamento de consultas. Na realização da consulta, o sistema consegue obter dados relevantes não só do ponto que recebeu a consulta, mas de qualquer outro conectado a este por algum caminho semântico [17].

No SPEED, qualquer ponto pode submeter uma consulta na rede utilizando apenas seu próprio esquema de dados, sem precisar conhecer esquemas de outros pontos. Dessa forma, a partir dos caminhos semânticos existentes, a consulta original pode ser reformulada aos outros pontos semânticos, de acordo com seus respectivos esquemas de dados [19].

O módulo responsável pela submissão e reformulação de consultas do sistema SPEED é chamado *SemRef* [17]. O objetivo deste trabalho é aperfeiçoar esse módulo, introduzindo um mecanismo de captura e armazenamento das preferências do usuário, a fim de serem utilizadas na personalização das consultas.

3.2 Módulo de Submissão de Consultas: *SemRef*

Conforme mencionado anteriormente, um dos principais serviços oferecidos pelo SPEED é o processamento de consultas. A consulta submetida em um ponto é analisada e reformulada a outros pontos de acordo com os seus esquemas. Ao final do processo, os resultados dos pontos que executaram a consulta são integrados naquele que iniciou a propagação para serem apresentados ao usuário.

O *SemRef* é o módulo responsável pela submissão e reformulação de consultas no SPEED. Ele identifica a semântica da consulta e, a partir de correspondências semânticas entre os pontos, efetua a sua reformulação para enviar a outros pontos da rede que possam respondê-la de maneira eficiente e satisfatória [17].

A semântica é uma das principais características que estão associadas ao *SemRef*, uma vez que esse conceito possibilita uma reformulação enriquecida e uma apresentação de resultados mais completos para as consultas. Entretanto, no processo de reformulação, um conceito presente em um determinado ponto da rede pode não apresentar equivalências em pontos de destino, dificultando, portanto, a obtenção de resultados exatos [23].

Para o desenvolvimento deste trabalho, o módulo *SemRef* considera unicamente dois pontos que desejam se comunicar através de uma consulta. Portanto, se um usuário submete uma consulta no ponto P1, ela será reformulada e encaminhada para apenas um ponto de destino P2.

A arquitetura geral do módulo de consulta *SemRef*, apresentada na Figura 6, destaca uma divisão em cinco componentes principais [22]: CODI, *Matcher* Semântico, Conjunto de Correspondências Semânticas, Manipulador da Consulta e Reformulador da Consulta. O CODI representa a ontologia de contexto responsável pelo armazenamento dos elementos contextuais associados com as consultas, entidades e operadores. O *Matcher* Semântico é responsável por gerar as correspondências semânticas existentes entre dois esquemas de

diferentes pontos da rede. O Conjunto de Correspondências Semânticas representa o mapeamento resultante do processo de *matching* semântico. O Manipulador da Consulta, além de ser responsável por receber os resultados e apresentá-los para o usuário, analisa a semântica da consulta, identificando as entidades e os operadores informados na *string* de busca do usuário. Por último, tem-se o Reformulador da Consulta, que através da verificação dos elementos contextuais e dos mapeamentos semânticos existentes, realiza a reformulação propriamente dita da consulta original.

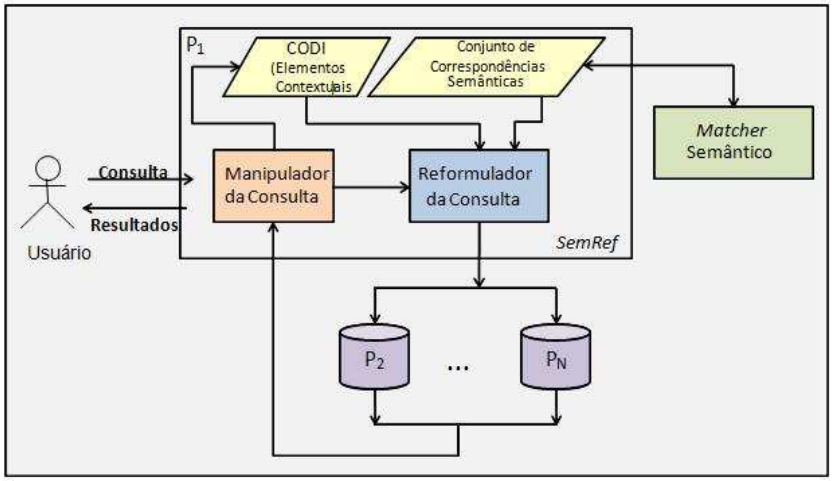

Figura 6 – Atual arquitetura do módulo *SemRef* [17]

3.3 Tratamento das Preferências do Usuário

O SPEED, como um típico sistema P2P, é configurado em um ambiente bastante dinâmico e flexível, onde cada ponto pode entrar e sair da rede a qualquer momento. O crescimento do número de participantes em uma rede pode dificultar a busca por informações relevantes aos interesses dos usuários. Com um universo de informações bastante variado, o tempo e esforço para encontrar resultados que abordam um tema específico crescem consideravelmente. A consulta poderá trazer um conjunto massivo de dados, muitas vezes, redundantes e irrelevantes. Além disso, diferentes usuários, ao submeterem uma determinada consulta, encontram sempre o mesmo conjunto de resultados na mesma ordem.

Dentro desse contexto, surge o desafio de fazer com que as consultas submetidas no SPEED retornem resultados que, da melhor forma possível, expressem as necessidades do

usuário. Para alcançar esse objetivo, este trabalho aborda a idéia de utilização das preferências dos usuários na personalização das buscas, realizando a filtragem da informação para organizar o volume dos dados que devem ser apresentados.

O desenvolvimento do mecanismo de utilização das preferências do usuário no SPEED foi refletida na arquitetura do *SemRef* com a adição de um novo componente: "Personalização das Consultas". Também se fez necessária a criação de uma base de dados para cadastrar os perfis dos usuários. A Figura 7 traz a arquitetura do *SemRef* com os componentes adicionados em destaque.

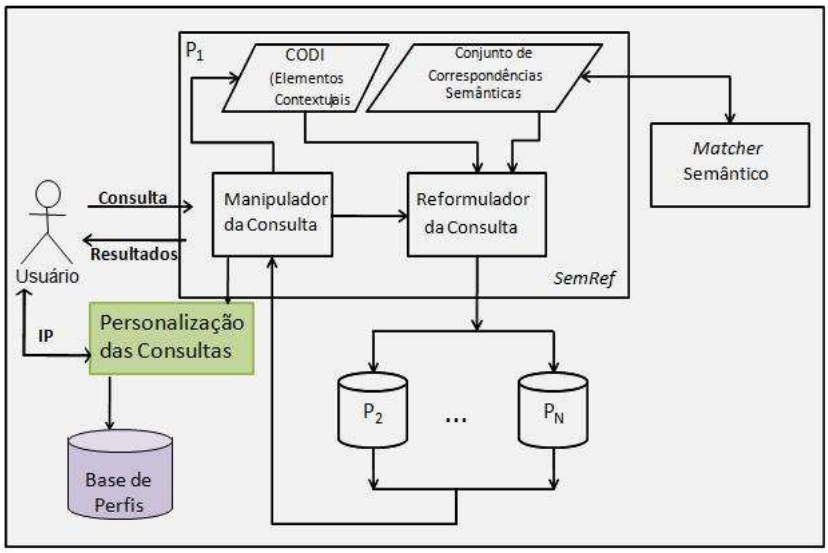

Figura 7 – Arquitetura do SemRef com o componente de Personalização

Conforme apresentado no Capítulo 2, existem diferentes abordagens para o tratamento da personalização de consultas. Neste trabalho, utiliza-se a técnica de personalização baseada em Perfil. Esta técnica faz uso das preferências e características pessoais dos usuários, relacionando-as com os esquemas de dados presentes nos pontos da rede. Portanto, quando o usuário submete uma consulta por um determinado conceito, serão relacionadas algumas propriedades desse conceito com a informação do perfil do usuário e, assim, as instâncias retornadas poderão ser ordenadas adequadamente.

Uma das formas de coletar características foi construída segundo a abordagem explícita, ou seja, requer a atuação direta do usuário para indicar as suas preferências.

Através do preenchimento de um formulário, o usuário define algumas informações pessoais e opiniões que poderão ser úteis na realização da consulta.

Para armazenar os perfis dos usuários foi utilizado um banco de dados relacional. Assim, ao utilizarem o *SemRef,* eles precisarão efetuar um *login* no sistema, de maneira que as suas informações sejam carregadas e utilizadas nas consultas.

No processo de modelagem de usuários foi definida uma estrutura que engloba informações como nome, idade, área de estudo, sexo e nacionalidade, além dos dados tradicionais de *login* e senha. Como tais informações possuem a tendência de permanecerem inalteradas ao longo do tempo de uso do sistema, pode-se classificar esse modelo como estático, conforme visto no Capítulo 2.

Essas contribuições objetivam a personalização das consultas no *SemRef,* re-classificando os resultados de acordo com as preferências do usuário, cujo perfil tenha sido carregado. Em uma primeira etapa, o processo original de consulta funciona da mesma maneira, reformulando e encaminhando a consulta para outros pontos. Apenas quando os resultados dos diferentes pontos forem integrados no ponto de submissão, a re-classificação ou ordenação será efetuada, sendo essa uma segunda etapa.

Considerando um domínio de Universidades como exemplo, quando um usuário, professor da disciplina de Banco de Dados, submeter uma consulta buscando por artigos ou publicações, é interessante que os primeiros resultados exibidos sejam aqueles que tratem assuntos relacionados à área de Banco de Dados. Da mesma forma, quando um professor de Engenharia de Software efetua a consulta, deseja encontrar primeiramente trabalhos voltados para a sua área.

Além da modelagem estática do usuário, constituída por informações que foram indicadas pelo próprio, houve a preocupação neste trabalho de definir um elemento contextual dinâmico: a localização do usuário. Para isso, no momento em que o usuário efetua o *login* no *SemRef,* é analisado o seu endereço IP com o objetivo de descobrir a partir de qual lugar o usuário irá submeter a consulta. Essa abordagem é particularmente interessante, uma vez que a estrutura de uma rede P2P pode envolver pontos localizados pelos mais diversos lugares no mundo.

Voltando para o exemplo no domínio de Universidades, quando um usuário do Canadá submete uma consulta buscando conferências ou palestras, é importante que sejam priorizados os resultados que tratem de conferências realizadas no próprio Canadá.

Conforme apresentado anteriormente, a extensão do módulo *SemRef* para tratamento das preferências e características dos usuários, definidas tanto estaticamente, através do perfil, quanto dinamicamente, através da localização por IP, originou novos casos de uso para o módulo de consultas do SPEED. A Figura 8 apresenta o diagrama dos casos de uso adicionados ao *SemRef*, obtidos a partir das contribuições deste trabalho.

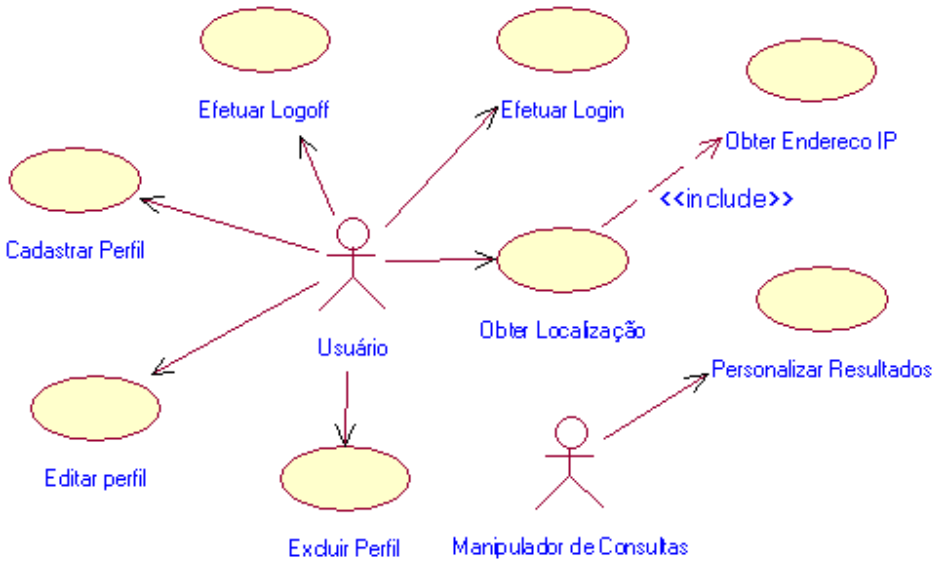

Figura 8 – Casos de usos adicionados ao *SemRef* para personalização

3.4 Considerações

Neste capítulo, foram apresentadas as modificações realizadas no módulo de consulta do SPEED, chamado *SemRef*, para introduzir funcionalidades que permitem o tratamento das preferências dos usuários, personalizando as consultas submetidas no sistema. A captura de características foi possível graças à utilização de um perfil do usuário, onde o mesmo indicava explicitamente as suas preferências. Além disso, também foi utilizada, como informação contextual dinâmica, a localização do usuário obtida a partir do endereço IP.

O capítulo seguinte descreve os detalhes de implementação dessas modificações, apresentando, inclusive, algumas telas de interface com o usuário.

4. Implementação

Neste capítulo serão apresentados alguns detalhes de implementação da estrutura de armazenamento e captura do perfil do usuário no SPEED. Primeiramente, serão descritas algumas ferramentas utilizadas no desenvolvimento do trabalho. Em seguida, uma visão geral das funcionalidades adicionadas será apresentada, através da exploração das telas de interface com o usuário. Por fim, será demonstrado um exemplo simplificado de como as consultas podem ser personalizadas no *SemRef*.

4.1 Ferramentas Utilizadas

Como o objetivo deste trabalho está voltado para a construção de uma extensão do módulo de consultas do SPEED, é necessária a utilização do mesmo ambiente de desenvolvimento utilizado para criação do *SemRef* originalmente. Diante disso, foi escolhida a linguagem Java [9], com o auxílio da ferramenta Eclipse [3]. Dentre os motivos que levaram a essa escolha está o fato da linguagem ser independente de plataforma e permitir a utilização de componentes em formas de API (*Application Program Interface*) para solucionar os mais diversos problemas. Além disso, a facilidade para construção de interfaces gráficas necessárias ao sistema também influenciou a escolha.

Para prover o armazenamento dos perfis dos usuários, foi utilizado como sistema de gerenciamento de dados o HSQLDB (*Hyperthreaded Structured Query Language Database*) [8]. Trata-se de uma solução escrita em Java bastante simples, que utiliza poucos recursos e possui um bom desempenho, principalmente em aplicações *desktops*.

Foi utilizada ainda outra biblioteca chamada GeoIP [15], distribuída pela empresa MaxMind. Através de uma versão gratuita dessa biblioteca, conseguiu-se obter dados referentes à localização do usuário, como o país, por exemplo, a partir da informação do endereço IP.

4.2 Telas de Interface com o Usuário

Conforme visto na Seção 3.2, o módulo de consulta do SPEED considera unicamente dois pontos que desejam se comunicar através de uma consulta. Sendo assim, a

interface com o usuário permite executar uma consulta em um ponto de origem, enquanto esta mesma consulta é reformulada a partir de correspondências semânticas entre o ponto de origem e outro de destino. Após essa reformulação, a consulta é executada no ponto de destino e os seus resultados são integrados para serem exibidos ao usuário. A Figura 9 ilustra a tela de consulta inicial do *SemRef*.

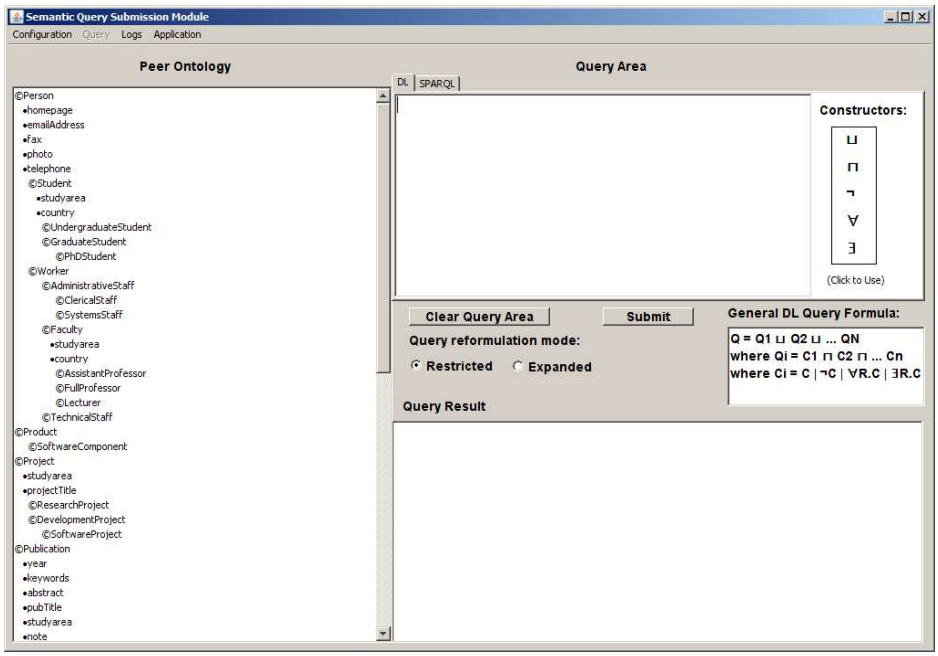

Figura 9 – Tela de consulta incial do *SemRef*

A interface de consulta apresenta no lado esquerdo a hierarquia dos conceitos que podem ser buscados no ponto, conforme o domínio escolhido. No lado superior direito fica área de formulação das consultas, que podem ser escritas tanto em lógica descritiva (DL, do inglês *Description Logic*), quanto em SPARQL. Os resultados das consultas serão exibidos no campo localizado na parte inferior direita da tela.

Devido o escopo deste trabalho ser relacionado com aspectos do tratamento de preferências do usuário e captura do perfil, não serão abordadas em maiores detalhes as funcionalidades da interface que fogem daquelas apresentadas pelos casos de usos definidos na Figura 8.

No *SemRef*, através do menu *Application* é possível ter acesso às principais funcionalidades adicionadas como forma de contribuição deste trabalho. A Figura 10 ilustra as opções do menu *Application*.

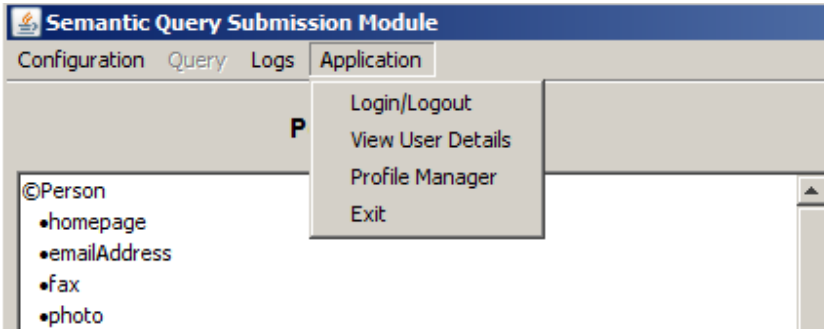

Figura 10 – Funcionalidades do menu *Application*

Conforme descrito no capítulo 3, para aplicação do conceito de personalização das consultas no SPEED, será necessário que o usuário indique as suas preferências. A coleta das características é realizada através do preenchimento de um formulário, disponível a partir do gerenciador de perfis (*Profile Manager*), cuja opção aparece no menu *Application*. Na Figura 11 é apresentada a interface do gerenciador de perfis.

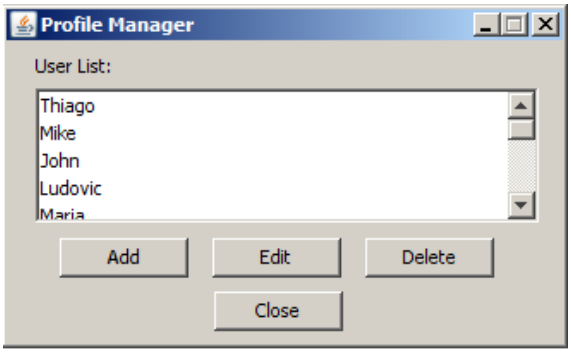

Figura 11 – Tela utilizada para gerenciar os perfis dos usuários

A partir da tela de gerenciamento de perfis, o usuário pode inserir, alterar ou excluir um perfil na base de dados do sistema. A Figura 12 traz o formulário utilizado pelo usuário para definir os seus dados pessoais ou preferências, conforme a modelagem explicada na Seção 3.3.

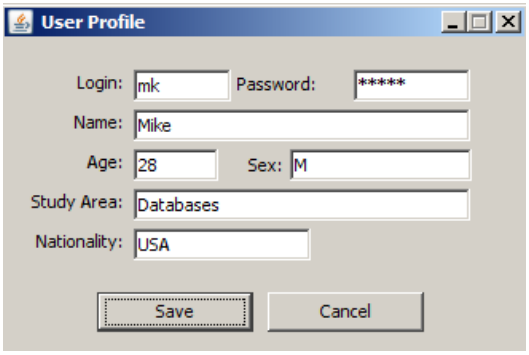

Figura 12 – Tela utilizada para adicionar ou alterar um perfil

Para que o módulo *SemRef* carregue o perfil e possa utilizá-lo na personalização das consultas, é necessário que os usuários se identifiquem no sistema. Essa identificação é realizada através da funcionalidade de *login*. A Figura 13 traz esta tela em destaque.

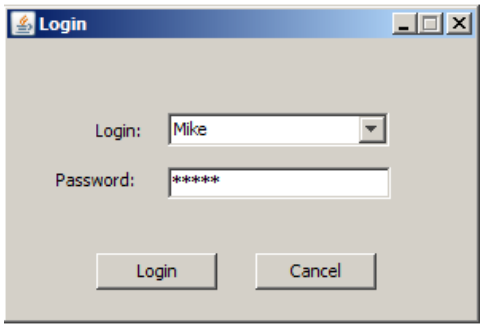

Figura 13 – Tela utilizada para efetuar *login*

A partir do momento em que o usuário realiza o *login* no sistema, são carregadas não apenas as informações do perfil estático, informado pelo o usuário através do gerenciador de perfil, mas também os elementos contextuais dinâmicos como o IP e a localização do usuário. Considerando que em uma rede P2P os usuários podem estar acessando o sistema de diferentes lugares no mundo, tais informações passam a ter uma importância significativa para a personalização da consulta. Quando se utiliza a funcionalidade de visualização dos dados do perfil do usuário (*View User Details*), disponível no menu *Application*, pode-se verificar o IP e o país correspondente encontrados pela aplicação. A Figura 14 traz a tela de visualização dos dados do perfil. É interessante

observar que a informação de nacionalidade fornecida pelo o usuário não é necessariamente equivalente à localização obtida através do IP, conforme ilustra a figura.

Figura 14 – Tela de informações do perfil do usuário após login

Por último, para o usuário encerrar as atividades com o sistema, ele deve acionar a opção *logout.* Quando se utiliza essa opção, a aplicação não levará em conta nenhuma informação de perfil do usuário, seja ela estática ou dinâmica. A Figura 15 apresenta a tela de *logout* do sistema.

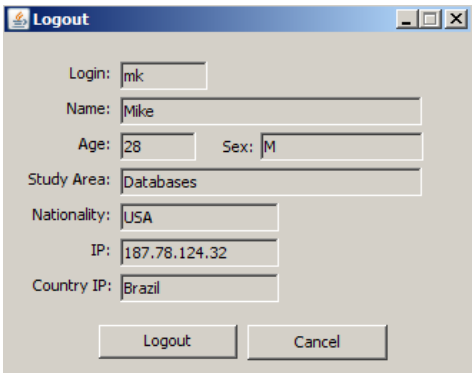

Figura 15 – Tela para efetuar *logout* da aplicação

4.3 Exemplo de consulta personalizada

Esta seção apresenta um exemplo simplificado de utilização das preferências do usuário na personalização de consultas no SPEED. Neste exemplo, dois usuários com características distintas realizam uma consulta por publicações científicas, através da

interface do *SemRef*. Vale ressaltar que o contexto considerado no exemplo é, mais uma vez, o domínio de Universidades.

Ambos os usuários possuem seus perfis cadastrados na base de dados do sistema. As Figuras 16 e 17 descrevem as informações de cada perfil. Observa-se que enquanto o primeiro usuário considera como área de estudo a Engenharia de Software, o segundo prefere estudar a área de Banco de Dados.

Figura 16 – Informações do perfil do primeiro usuário

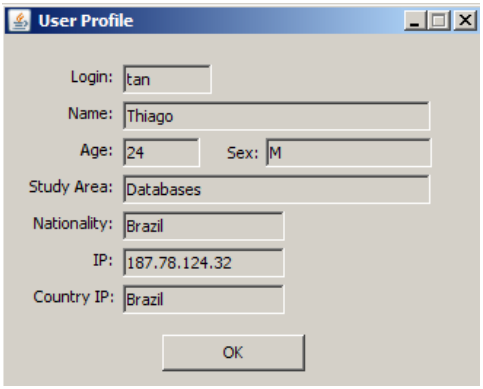

Figura 17 – Informações do perfil do segundo usuário

No momento da realização da consulta, as preferências e características de cada usuário serão relacionadas com o esquema de dados tanto do ponto de origem, onde a consulta foi submetida, quanto no ponto de destino. Assim, as propriedades relacionadas com o conceito de Publicação serão comparadas com as informações do perfil do usuário, especialmente os dados referentes ao campo de estudo.

Os relacionamentos e comparações efetuadas são refletidos na ordenação das instâncias retornadas pela consulta. Conforme é observado na Figura 18, os resultados de

publicações com tema associado à área de Engenharia de Software são destacados na consulta efetuada pelo primeiro usuário. Já na consulta submetida pelo segundo usuário, os resultados apresentados inicialmente são aquelas publicações relacionadas à área de Banco de Dados. A Figura 19 exibe os resultados retornados por essa última consulta.

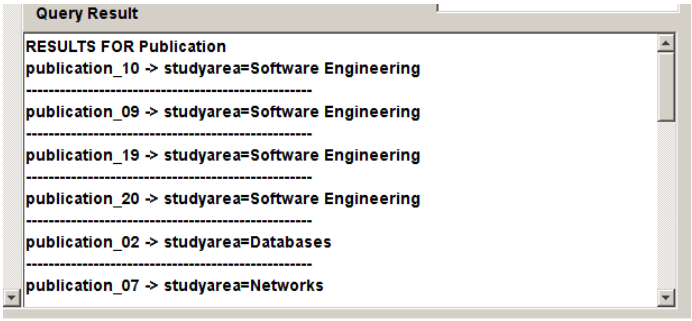

Figura 18 – Resultado priorizando publicações sobre Engenharia de Software

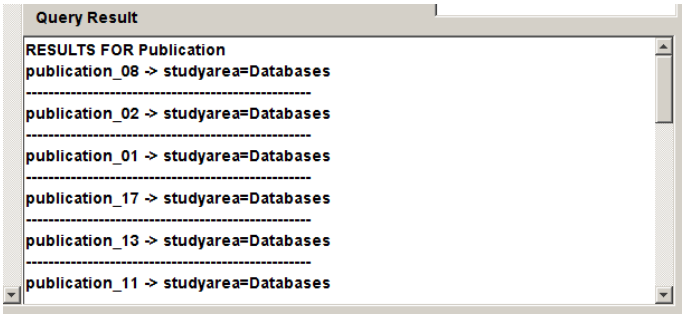

Figura 19 – Resultado priorizando publicações sobre Banco de Dados

4.4 Considerações

Este capítulo abordou os detalhes de implementação da estrutura para personalização das consultas no SPEED. Foram apresentadas as bibliotecas utilizadas e as telas de interface para as funcionalidades de armazenamento e captura do perfil do usuário. Além disso, foi mostrado um exemplo simplificado de submissão de consulta no sistema, utilizando diferentes configurações de perfil.

No próximo capítulo, serão descritas as conclusões obtidas a partir desse estudo, bem como sugestões para trabalhos futuros.

5. Conclusão

Este capítulo apresenta algumas considerações finais a respeito deste trabalho, especificando seus objetivos principais. Em seguida, são levantadas algumas sugestões de trabalhos futuros, visando o aperfeiçoamento da solução proposta.

5.1 Considerações Finais

Este trabalho destacou a importância do tratamento das preferências do usuário na composição dos sistemas de informação atuais. A sobrecarga de informações em alguns ambientes, especialmente a web, transformou a simples tarefa de realizar uma pesquisa em uma atividade dispendiosa. A grande maioria das instâncias apresentadas, como resultado de uma consulta, não corresponde ao que de fato o usuário necessita. Por isso, foi enfatizado nesse estudo que, com base nas preferências, é possível personalizar as consultas e reduzir os esforços empregados para localização de conteúdo relevante.

Em ambientes do tipo *peer-to-peer* a dificuldade na submissão de consultas não foge à regra. Os sistemas de gerenciamento de dados baseados nessa topologia, como o SPEED, costumam trazer como resultado um conjunto massivo de dados, muitas vezes, desnecessários.

Com base nessa questão, o trabalho propôs modificações no funcionamento do módulo de consultas do sistema SPEED. Foi introduzido neste módulo um mecanismo de armazenamento e captura do perfil do usuário, de maneira que as características do mesmo passaram a ser utilizadas na personalização das consultas.

Inicialmente, foram definidos neste trabalho alguns conceitos relacionados com preferências do usuário, personalização de consultas e modelagem de usuários. A apresentação desses conceitos foi realizada em conjunto com uma revisão da literatura nessas áreas. Em seguida, foram introduzidos alguns fundamentos relacionados com o SPEED, juntamente com a descrição das contribuições deste trabalho no módulo de submissão de consultas. Por último, foram apresentados alguns detalhes da implementação da estrutura de personalização desenvolvida.

5.2 Trabalhos Futuros

As contribuições deste trabalho consistem na estrutura de tratamento das preferências dos usuários adicionada no módulo de consulta do SPEED. A coleta das características é realizada de forma explícita, quando o próprio usuário indica as suas preferências, ou de forma implícita, quando se utiliza a informação contextual dinâmica de localização a partir do IP.

Além da abordagem de personalização baseada em perfil, vista neste trabalho, o módulo de consulta pode ser incrementado através da utilização de outras categorias de personalização, tais como baseada em comportamento ou baseada em colaboração.

De maneira semelhante como Micarelli [16] propõe em alguns de seus trabalhos, a personalização baseada em comportamento, no SPEED poderia ser realizada através da manutenção do histórico das consultas realizadas pelos usuários. Dessa forma, as consultas podem ser adaptadas a partir de resultados obtidos anteriormente. Já a personalização baseada em colaboração, através de conceitos como votação e opinião, pode influenciar a consulta não apenas a partir de suas preferências pessoais, mas também das preferências de outros usuários com perfil semelhante.

Referências Bibliográficas

[1] Alves, C. R. C.; Filgueiras, L. V. L. Avaliação Comparativa de Algoritmos de Personalização para Direcionamento de Conteúdo, In: Second Latin American Conference on Human-Computer Interaction, CLIHC, 2005.

[2] Chomicki, J. Querying with Intrinsic Preferences. In: 8th International Conference on Extending Database Technology: Advances in Database Technology, p.34-51. Mar. 2002.

[3] Eclipse IDE. Disponível em: < http://www.eclipse.org/>. Acesso em: 26 jun. 2010.

[4] Eurekster. Disponível em: <http://www.eurekster.com>. Acesso em: 26 jun. 2010.

[5] Ferreira, Aurélio Buarque de Holanda. Dicionário Aurélio Básico da Língua Portuguesa. Rio de Janeiro: Nova Fronteira, 1988.

[6] Giannopoulos G., Dalamagas T., Sellis T. Collaborative Ranking Function for Web Search Personalization. In: Proceedings of the 3rd International Workshop on Personalized Access, Profile Management and Context Awareness in Databases (PersDB 2009), Lyon, France.

[7] Google Search History. Disponível em: < http://www.google.com/searchhistory>. Acesso em: 26 jun. 2010.

[8] HSQLDB. Hyperthreaded Structured Query Language Database. Disponível em: <http://hsqldb.org/>. Acesso em: 26 jun. 2010.

[9] Java Platform SE 6. Disponível em: <http://java.sun.com/javase/6/docs/api/>. Acesso em: 26 jun. 2010.

[10] Koutrika G., Ioannidis Y. Personalized Queries under a Generalized Preference Model. In: 21st Intl. Conf. On Data Engineering (ICDE), p.841-852, Tokyo, Japan, 2005.

[11] Koutrika, G. Personalization of Structured Queries with Personal and Collaborative Preferences. In: ECAI Workshop about Advances on Preference Handling, Riva del Garda, Italy. 2006.

[12] Koutrika, G. Query Personalization based on User Preferences. Disponível em <http://cgi.di.uoa.gr/~phdsbook/files/OK_Koutrika.pdf>. Acesso em: 17 mar. 2010.

[13] Marchi, K. R. C. Uma Abordagem para Personalização de Resultados de Busca na Web. Universidade Estadual de Maringá. Maringá, 2010.

[14] Marchi, K. R. da C. Personalização De Consultas Web: Uma Abordagem Geral. Universidade Estadual de Maringá - UEM. Maringá.

[15] MaxMind GeoIP. Disponível em: <http://www.maxmind.com/app/geolitecountry>. Acesso em: 26 jun. 2010.

[16] Micarelli, A; Gasparetti, F.; Sciarrone, F.; GAUCH, S. Personalized Search on the World Wide Web. In The Adaptive Web. Alemanha. Maio, 2007.

[17] Neves T. Desenvolvimento do Módulo de Reformulação de Consultas no Sistema SPEED. Monografia de Conclusão de Curso. Universidade Federal de Pernambuco (UFPE), Recife, PE, Brasil, 2008.

[18] Oliveira, E. W. Personalização: Um Estudo no Contexto de Aprendizagem. Universidade Federal do Estado do Rio de Janeiro (UNIRIO), Rio de Janeiro, Julho de 2007.

[19] Pires, C.E."Um Sistema P2P de Gerenciamento de Dados com Conectividade Baseada em Semântica". Monografia de Qualificação e Proposta de Doutorado, CIn, UFPE. Abril, 2007

[20] Santos, C. T. and Osório, F. S. 2004. An intelligent and adaptive virtual environment and its application in distance learning. In Proceedings of the Working Conference on Advanced Visual interfaces (Gallipoli, Italy, May 25 - 28, 2004).

[21] Silva, L. A. M. OWLPREF: Uma Representação Declarativa de Preferências Qualitativas para a Web Semântica. Universidade de Fortaleza. Fortaleza, 2007.

[22] Souza D. Using Semantics to Enhance Query Reformulation in Dynamic Distributed Environments. PhD Thesis, Federal University of Pernambuco (UFPE), Recife, PE, Brazil, 2009.

[23] Souza D., Neves T., Salgado A. C., Tedesco P., Kedad Z. Semantic-based Query Reformulation in Dynamic Environments. In: 25e Journées Bases de Données Avancées (BDA'09), Namur, Belgique, 2009.

[24] SPEED. Semantic Peer-to-Peer Data Management System. Disponível em <http://www.cin.ufpe.br/~speed> Acesso em: 26 mar. 2010.

[25] Stefanidis K., Pitoura E., Vassiliadis P. On Supporting Context-Aware Preferences in Relational Database Systems. In Proc. of the first International Workshop on Managing Context Information in Mobile and Pervasive Environments (MCMP'2005), in conjunction with MDM 2005, Cyprus. 2005.

[26] Sung L. G. A., Ahmed N., Blanco R., Li H., Soliman M. A., Hadaller D. (2005): A Survey of Data Management in Peer-to-Peer Systems. School of Computer Science, University of Waterloo.

[27] WIKIPEDIA. Preference. Disponível em: <http://en.wikipedia.org/wiki/Preference>. Acesso em: 26 jun. 2010.