

SINCRONIZANDO BASES DE DADOS EM AMBIENTES DISTRIBUÍDOS: UM ESTUDO DE CASO NO PROJETO SIEP GERENCIAL

Jackson Amaral da SILVA; Márcio Rodrigo Melo MARTINS; Rafael Fernandes LOPES; Omar Andrés Carmona CORTES; Dejailson Nascimento PINHEIRO

Centro Federal de Educação Tecnológica do Maranhão (CEFET – MA). Avenida Getúlio Vargas, N°. 04, Monte Castelo, CEP: 65020-300. E-mail: jackson.amarals@gmail.com, marciorodrigomm@gmail.com, rafaelf@cefet-ma.br, omar@cefet-ma.br, dejailson.pinheiro@gmail.com

RESUMO

A grande quantidade de sistemas de informação existentes nas instituições de EPTs dificulta enormemente o acesso a essa informações pelo MEC. Com o objetivo de minorar esta dificuldade, a SETEC propôs o desenvolvimento de uma infra-estrutura de software chamada SIEP Gerencial, cujo principal objetivo é integrar as diversas bases de dados de EPT existentes no País em uma única base de dados confiável e atualizada. Esta infra-estrutura provê mecanismos de atualização contínua de dados e de acesso a informações por usuários e sistemas computacionais. O SIEP Gerencial está sendo atualmente desenvolvido por alunos e professores da própria rede de EPT de forma colaborativa. Este artigo trata sobre o módulo Sincronizador, um dos sub-sistemas que compõem o SIEP Gerencial. O Sincronizador tem o intuito de manter as bases de dados dos sistemas utilizados pela SETEC/MEC sincronizadas com a base nacional do SIEP. Além disso, este módulo provê, entre suas funcionalidades: (a) o mapeamento entre bases de dados através de uma interface gráfica e (b) a transferência de dados de forma manual ou automática (por meio de agendamentos). Toda a comunicação entre o módulo Sincronizador e os demais sub-sistemas do SIEP Gerencial é feita por meio da tecnologia de serviços web.

Palavras-chave: Sistemas de Informação, Educação Profissional e Tecnológica, SIEP Gerencial

1. INTRODUÇÃO

A grande expansão experimentada pela rede de Educação Profissional e Tecnológica (EPT) no Brasil tem motivado o desenvolvimento de uma grande quantidade de sistemas de informação, tanto nas unidades acadêmicas quanto nas instâncias de regulação e avaliação. Entretanto, estes dados encontram-se dispersos através desta rede, não permitindo que a Secretaria de Educação Profissional e Tecnológica (SETEC/MEC) tenha acesso a informações, estatísticas e indicadores fidedignos sobre a EPT no Brasil. A falta destas informações dificulta o papel da SETEC, enquanto definidora de políticas públicas de EPT.

Nesse contexto, nasceu o projeto SIEP Gerencial, cujo objetivo é prover um ambiente que permita ao MEC melhorar os processos de planejamento estratégico e operacional, bem como suas rotinas administrativas, acadêmicas e de gestão, estabelecendo os indicadores necessários ao diagnóstico, monitoramento e avaliação das EPTs no país (MANFREDI, 2003). Enxergando o lado da instituição, faltam instrumentos e ferramentas que possibilitem sua gestão efetiva, tanto acadêmica quanto administrativa, garantindo a integração das bases de dados locais a cada instituição de ensino com a SETEC/MEC.

O SIEP Gerencial proverá à SETEC/MEC instrumentos e ferramentas que possibilitem o exercício de sua função definidora de políticas, supervisora, estimulando um processo contínuo de avaliação, monitoramento, modernização e transparência da oferta e da expansão da Educação Profissional e Tecnológica no Brasil. Este projeto é desenvolvido de forma colaborativa por alunos e professores da rede federal de EPT, fortalecendo as áreas de pesquisa e inovação, ensino, gestão e extensão dentro desta própria rede.

A infra-estrutura do SIEP Gerencial é responsável por coletar informações diretamente das bases acadêmicas de toda a rede de EPT e mantê-las atualizadas em uma base de dados nacional. Assim, faz-se necessária a existência de um mecanismo que permita que estes dados atualizados sejam propagados para diversas outras bases de dados (heterogêneas) existentes no MEC, de forma a mantê-las atualizadas. Este mecanismo foi implementado como um módulo do projeto SIEP Gerencial, e é chamado de **Sincronizador**. O principal objetivo do módulo Sincronizador é manter atualizadas as diversas bases de dados de "sistemas-alvos" (i.e. os diversos sistemas computacionais já existentes no MEC que necessitam de atualizações de dados sobre a EPT no Brasil).

Este artigo esta dividido da seguinte forma: a Seção 2 apresenta o projeto SIEP e sua arquitetura; a Seção 3 mostra detalhes de implementação do módulo Sincronizador; a Seção 4 traz informações a respeito das tecnologias utilizadas no desenvolvimento do módulo Sincronizador; e a Seção 5 apresenta as conclusões deste trabalho.

2. O PROJETO SIEP GERENCIAL

O projeto SIEP Gerencial pretende suprir a necessidade do MEC de coletar informações de maneira eficiente. Em outras palavras, o SIEP-G irá centralizar essas informações usando formatação padrão de dados. O SIEP Gerencial atua desde a extração dos dados a partir dos sistemas de informação de cada unidade acadêmica, até a disponibilização dessas informações para os vários sistemas no âmbito do Ministério da Educação (ALEIXO et al., 2007).

A arquitetura do SIEP Gerencial é formada por cinco componentes (Extrator, Atualizador, Seletor, Relator e Sincronizador) conforme ilustrado na Figura 1. Embora os componentes sejam independentes eles são complementares, pois alguns módulos dependem de outros.

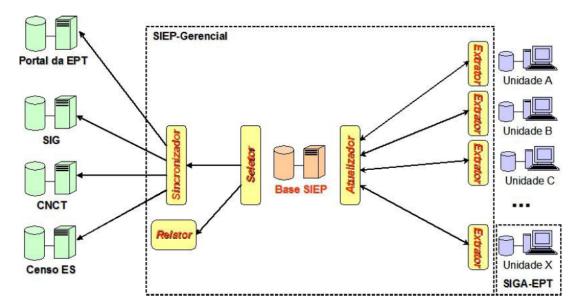


Figura 1 - Arquitetura do SIEP Gerencial

Os módulos Seletor, Atualizador e Extrator foram desenvolvidos com base na filosofia SOA (*Service Oriented Architecture* – Arquitetura Orientada a Serviços), sendo implementados através de serviços web. Esse fraco acoplamento proporcionado pela arquitetura escolhida irá facilitar a manutenção e as futuras alterações (WOODS, 2006). O Sincronizador, alvo deste trabalho, foi desenvolvido como uma aplicação *stand-alone*. A seguir são apresentadas as funções de cada módulo.

- Extrator Este módulo tem como finalidade a extração dos dados necessários ao SIEP Gerencial a partir das várias bases de dados utilizadas pelas unidades acadêmicas de EPT. Para a coleta das informações extraídas, o módulo Extrator disponibiliza um serviço web para receber a solicitação de coleta (HANSEN, 2007). O módulo Extrator permite também que o gestor da unidade acadêmica de EPT visualize os dados extraídos. Dessa forma, é possível verificar os dados antes de realizar alterações e por fim autorizar a coleta desses dados.
- Atualizador É responsável pela coleta dos dados a partir dos Módulos Extratores, instalados em cada unidade acadêmica de EPT, e atualizar a base de informações central do SIEP Gerencial. A partir de um escalonamento definido pelo operador do sistema, o módulo Atualizador aciona o serviço web do módulo Extrator do qual deseja receber os dados extraídos. Para receber os dados extraídos, o módulo Atualizador disponibiliza um serviço web para receber a coleta. Tão logo receba o resultado de uma coleta, é iniciado um processo de atualização da base de informações central do SIEP (ALEIXO et al, 2007).
- Seletor Este módulo tem a função inversa do Atualizador, ou seja, seu objetivo é fornecer acesso à base do SIEP Gerencial através de um serviço web. Essas informações podem ser posteriormente utilizadas tanto pelo módulo Sincronizador quanto pelo módulo Relator. O Seletor fornece basicamente serviços para obter metadados e processar consultas em SQL e XQL (uma linguagem de consulta baseada em XML). Para evitar problemas com diferentes padrões de SQL (relativos a cada SGBD), em um futuro próximo, o Seletor processará apenas consultas em XQL, as quais poderão ser geradas a partir de um editor gráfico de consultas.
- Relator O módulo Relator visa fornecer acesso aos dados da base SIEP na forma de relatórios. Os relatórios podem ser formatados e criados de acordo com as necessidades do usuário, isto é, o usuário pode selecionar os dados que deseja visualizar ou sumarizar e formatá-los do modo que mais lhe convém. O acesso à base de dados é feita através do módulo seletor utilizando consultas em XQL, sendo que o relator é uma aplicação Web.
- Sincronizador A função deste módulo é manter as bases dos sistemas alvos sincronizadas com a base do SIEP. Assim como o relator, para poder acessar os dados este módulo precisa solicitar as informações ao módulo seletor. Atualmente o Sincronizador envia consultas ao Seletor no formato SQL ANSI. Além disso, o modulo Sincronizador possui um mecanismo de agendamento de sincronização, permitindo que sincronizações sejam feitas sem a intervenção direta de um usuário. Detalhes sobre o módulo Sincronizador podem ser visto na próxima seção.

3. O SINCRONIZADOR

O Sincronizador tem o objetivo de manter atualizadas as bases de dados dos diversos Sistemas-alvo existentes no MEC, com todas as informações da Educação Profissional e Tecnológica do país. Através deste mecanismo é possível que estes sistemas, mesmo que já legados, continuem fornecendo serviços ao Ministério com base em informações fidedignas e atuais.

Para realizar esta tarefa o Sincronizador comunica-se periodicamente com o módulo Seletor, com vistas a solicitar informações atualizadas de EPT e transferir para as bases de dados de destino (ou mesmo atualizar estes dados). A Figura 2 apresenta a arquitetura dos módulos (componentes) de software que compõem o Sincronizador.

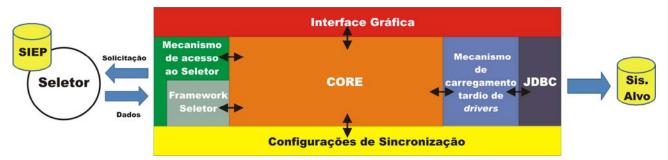


Figura 2 – Arquitetura do módulo Sincronizador

Como pode ser observado em sua arquitetura, o Sincronizador deverá ter duas camadas de comunicação: (a) a primeira, de acesso ao Seletor, que utilizará a tecnologia de serviços web para obter informações da base SIEP a partir de consultas pré-configuradas e (b) a segunda, de acesso à base dos diversos Sistemas-alvo do MEC, que utilizará a tecnologia JDBC. A camada de acesso ao Seletor envia solicitações contendo consultas no formato da linguagem SQL ou XQL. A partir destas solicitações o Seletor responderá (de forma síncrona ou assíncrona) contendo as informações requisitadas em formato XML. A manipulação das informações coletadas a partir do Seletor será realizada através do framework Seletor. Todo este ciclo de comunicação será controlado pelo *core* (núcleo) do sistema. É esta camada que controla os fluxos de informações entre a base SIEP e as bases dos Sistemas-alvo, implementando as regras de negócio do processo de sincronização. Este processo é configurado a partir de uma interface gráfica intuitiva, através da qual é possível gerar os mapeamentos, agendar e executar as sincronizações. Todos os dados necessários à execução das Sincronizações ficam armazenados como configurações de sincronização. Este ciclo de comunicação pode ser visto na Figura 3.

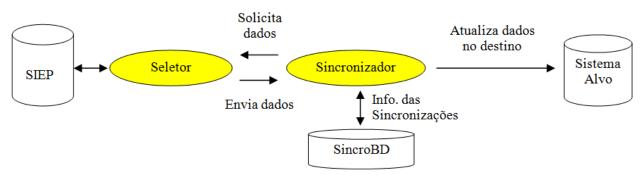
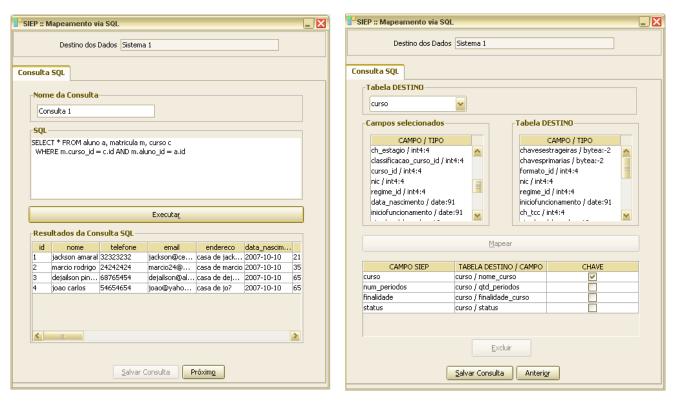


Figura 3. Comunicação entre o Sincronizador, o Seletor e a base do Sistema-alvo

O Sincronizador atua como um mediador das transferências de dados entre a base SIEP e a base do Sistema-alvo, transportando os dados de forma que estes não se dupliquem, mas permaneçam atualizados na base do sistema-alvo. Deve-se destacar que a sincronização só pode ser feita após a configuração do mapeamento entre a base SIEP e a base de destino (isto será feito por um operador credenciado pelo MEC e conhecedor da base de dados do sistema-alvo). Durante este mapeamento, o operador do sistema precisa indicar pelo menos um campo que representará a chave de sincronização. Esta informação é essencial para o processo, uma vez que o Sincronizador verifica se a mesma chave existe na base SIEP e na base do sistema-alvo. Se a mesma chave existir, os dados dos demais campos do registro são atualizados. Entretanto, se a chave de um dado registro da base SIEP não existir na base do sistema-alvo, este registro é completamente transferido

para o destino. As telas do Sincronizador que ilustram o processo de mapeamento entre estas bases de dados são apresentadas nas Figuras 4(a) e (b).



(a) Geração de consulta à base SIEP

(b) Mapeamento da consulta gerada com a base destino

Figura 4. Processo de mapeamento entre a base SIEP e as bases dos sistemas-alvo do MEC

Para que o Sincronizador possa inserir ou atualizar os dados recebidos, é necessário registrar uma base de dados. Para prover mais flexibilidade ao sistema o sincronizador permite registrar diferentes tipos de bases de dados (já que os diversos Sistemas-alvo diferem em vários aspectos – inclusive no SGBD utilizado). Os drivers de acesso a estes SGBDs (previamente registrados) são carregados a partir de um mecanismo de "carregamento tardio", que permite que o Sincronizador os carregue de forma dinâmica e em tempo de execução, utilizando recursos de reflexão computacional da linguagem Java. Um mecanismo de Agendamento de Sincronizações torna esta solução ainda mais eficaz, uma vez que permite ao operador do MEC informar a data, hora e a periodicidade da execução destas sincronizações. A interface gráfica de cadastramento destes agendamentos pode ser vista na Figura 5.

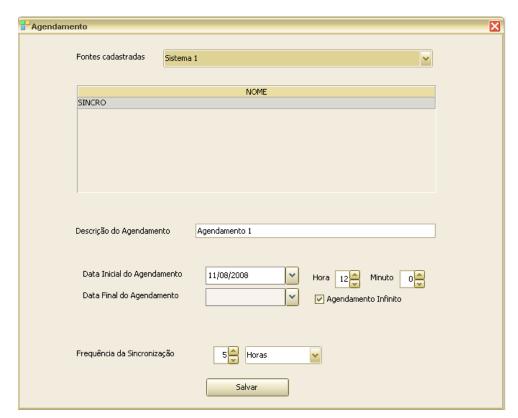


Figura 5 – Interface gráfica de cadastramento de agendamentos

Antes de executar o Sincronizador pela primeira vez é imprescindível informar em seus arquivos de configuração a URL do serviço seletor e as informações sobre sua conexão com o banco de dados. Isto se faz necessário pois no momento da inicialização o sincronizador tenta estabelecer uma conexão com seu banco de dados e com o serviço seletor. Se tiver sucesso o sistema inicia normalmente, caso contrário o sistema não inicia.

O Sincronizador é dividido basicamente em cinco módulos como pode ser observado no diagrama de domínio na Figura 6. Apesar do diagrama representar como as classes se relacionam entre si, o termo "módulos" é utilizado de forma genérica, assim, é possível entender como os módulos foram implementados e como eles se relacionam.

O módulo de Conexões (Conexão) permite que o usuário cadastre e gerencie diversos SGBDs que a partir de então poderão ser acessados pelo sistema, para isso é necessário apenas especificar um arquivo com as informações da conexão e o Driver JDBC para o SGBD.

O módulo de sistemas alvo (Fonte) permite que o usuário cadastre e gerencie os sistemas-alvo que poderão ser acessados pelo sincronizador, selecionando o SGBD que ele acessa e informando dados sobre a conexão.

O módulo de mapeamentos (Mapeamento) consiste basicamente no processamento de um script SQL de seleção (a ser executado na base do SIEP pelo serviço seletor) e um conjunto de associações entre campos das tabelas da base SIEP e campos das tabelas da base do sistema-alvo selecionado. Cada associação consiste no nome de um campo de uma tabela SIEP (campo de origem), o nome de uma tabela da base do sistema-alvo (tabela de destino) e o nome de um campo dessa tabela (campo de destino). Para realizar o cadastro o usuário seleciona o sistema-alvo que deseja sincronizar, e em seguida, por meio de um script SQL, seleciona os dados que deseja extrair da base do SIEP. Os resultados então são visualizados e depois se associam os campos selecionados aos campos de alguma tabela do sistema alvo. Essa associação é feita graficamente, observando que para cada tabela do sistema alvo envolvida no mapeamento é necessário definir uma associação que exercerá o papel de chave primária do mapeamento naquela tabela. Além disso, o sistema ainda impede que seja realizada uma associação entre dois campos com tipos de dados incompatíveis. As interfaces utilizadas para a realização deste processo foram apresentadas nas Figuras 4(a) e (b).

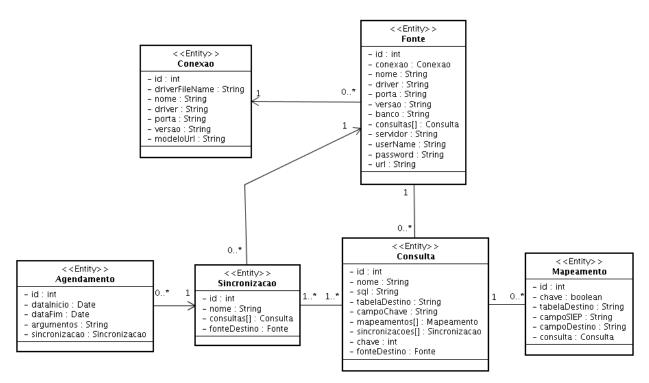


Figura 6 - Diagrama de domínio do Sincronizador

O módulo de sincronizações é subdividido em outros dois sub-módulos (Sincronização e Consulta). O primeiro é o módulo de cadastro e gerenciamento de sincronizações, que permite que o usuário adicione sincronizações para um sistema-alvo selecionado. Uma sincronização consiste de uma lista ordenada de mapeamentos, que são selecionados dentre os mapeamentos cadastrados e identificados por um nome (que é o nome da sincronização). O segundo sub-módulo é o módulo de execução de sincronizações, que permite que o usuário selecione sincronizações e as executem.

A execução das sincronizações consiste em executar o script SQL de cada mapeamento na base SIEP, por meio do serviço seletor, e em seguida inserir cada campo do resultado na base do sistema-alvo de acordo com as associações de tabela de destino e campo de destino, observando os valores dos campos definidos como chaves dos mapeamentos, executando em primeiro lugar um comando UPDATE. Caso o número de linhas afetadas seja igual a zero é executado então um comando INSERT, garantindo assim que os registros que já existem sejam atualizados e os que não existem sejam inseridos. É importante ressaltar que a ordem em que os dados são transferidos é a mesma ordem que os mapeamentos se encontram dentro da sincronização, essa ordem é importante para que as tabelas que possuem chaves estrangeiras sejam sincronizadas somente após a sincronização das tabelas que elas referenciam, mantendo assim a integridade referencial.

O Agendamento permite que o processo de sincronização das bases de dados seja mais ágil, pois permite o agendamentos de sincronizações executadas periodicamente. Este modulo é especificamente um cadastro de agendamentos que torna o processo de sincronizações automatizados, pois as informações são salvas na base de dados do próprio sincronizador.

4. TECNOLOGIAS UTILIZADAS NO SINCRONIZADOR

O Sincronizador foi desenvolvido em linguagem Java (ARNOLD et. al, 2000) na sua versão JEE (*Java Enterprise Edition*), que compreende um vasto conjunto de tecnologias que abrangem uma série de especificações voltadas principalmente para aplicações corporativas, contendo quase todos os pontos vitais para o desenvolvimento de aplicações de alta disponibilidade (MCLAUGHLIN, 2002; NICHOLAS, 2000; SAMPAIO, 2007).

A comunicação entre o sincronizador e o seletor é feita utilizando-se serviços web. Um serviço web é uma solução utilizada na integração de sistemas e na comunicação entre aplicações diferentes. Com esta tecnologia é possível que novas aplicações possam interagir com aquelas que já existem e que sistemas desenvolvidos em plataformas diferentes sejam compatíveis. Além disso, os serviços web são componentes

que permitem às aplicações enviar e receber dados em formato XML. Cada aplicação pode ter a sua própria "linguagem", que é traduzida para uma linguagem universal, o formato XML.

Os módulos que precisam de persistência utilizam uma tecnologia denominada *Java Persistence API* (JPA) (KEITH, 2006; BURKE, 2006), um framework para o mapeamento objeto-relacional escrito na linguagem Java. O objetivo da JPA é diminuir a complexidade entre os programas Java, baseado no modelo orientado a objeto, que precisam trabalhar com um banco de dados do modelo relacional (presente na maioria dos SGDBs). Em especial, no desenvolvimento de consultas e atualizações dos dados.

A principal característica do JPA é a transformação das classes em Java para tabelas de dados (e dos tipos de dados Java para os da SQL). A JPA gera as chamadas SQL e libera o desenvolvedor do trabalho manual da conversão dos dados resultante, mantendo o programa portável para quaisquer bancos de dados SQL, porém causando um pequeno aumento no tempo de execução. No sincronizador as informações persistentes são armazenadas, inicialmente, em um banco de dados PostgreSQL.

Para o desenvolvimento do módulo agendador foi utilizado um *framework* denominado Quartz, que pode ser definido como um agendador de tarefas. Portanto, o Quartz é um *framework* que tem como responsabilidade notificar outros componentes de software quando um pré-determinado horário ocorre. Nesse contexto, tornase possível executar tarefas de forma agendada, sejam essas tarefas periódicas ou não.

5. CONCLUSÕES

Este artigo apresentou o desenvolvimento de um sistema sincronizador cujo propósito é manter bases de dados dos diversos sistemas-alvo existentes no MEC atualizados em relação aos dados reais da rede. Este é um importante recurso do projeto SIEP Gerencial, uma vez que permite que uma série de sistemas (mesmo os legados) que se encontram em utilização, continuem exercendo seus papéis de forma eficaz, contando agora com informações fidedignas sobre a rede nacional de EPT. Por conta da utilização do Sincronizador, diversos sistemas que estariam fadados ao desaparecimento, agora poderão continuar em operação, gerando uma valorização do capital investido em desenvolvimento, implantação e treinamento nestes sistemas.

Através do projeto SIEP Gerencial será possível à SETEC/MEC manter um panorama completo da EPT nacional. Assim será possível que a mesma cumpra o seu papel principal de definição de políticas públicas. Outra importante contribuição que será alcançada com o SIEP Gerencial é o desenvolvimento de um modelo de desenvolvimento de sistemas de forma cooperada e em rede, já que estão participando efetivamente do desenvolvimento do sistema, 8 (oito) Centros Federais de Educação Profissional e Tecnológica, sendo eles: CEFET-AM, CEFET-MA, CEFET-MT, CEFET-PB, CEFET-PE, CEFET-Química-RJ, CEFET-RN e CEFET-SC.

Por outro lado, este trabalho apresenta uma contribuição tecnológica na área de banco de dados, mas precisamente, este trabalho apresentou uma nova abordagem de sincronização em bancos de dados distribuídos utilizando soluções de software livre.

REFERÊNCIAS

ALEIXO, Fellipe Araújo; AZEVEDO da Silva, Gilbert; dos SANTOS, Adriano Bezerra; RODRIGUES, Thyago Barbosa; do CARMO, Michelle Furtado Pinheiro, SIEP GERENCIAL: SISTEMA DE CENTRALIZAÇÃO DAS INFORMAÇÕES DA EDUCAÇÃO PROFISSIONAL E TECNOLÓGICA, II Congresso de Pesquisa e Inovação da Rede Norte Nordeste de Educação Tecnológica, João Pessoa – PB, 2007.

ARNOLD, Ken, GOSLING, James, HOLMES, David, The Java Programming Language, third edition, Addison-Welsley, 2000.

BURKE, Bill; MONSON-HAEFEL, Richard. Enterprise JavaBeans 3.0. 5. ed. Sebastopol, Estados Unidos: O'Reilly, 2006.

HANSEN, Mark. D. SOA Using Java Web Services. Prentice Hall, 2007.

KEITH, Michael; HALEY, Jason; SCHINCARIOL, Merrick. PRO EJB 3 – Java Persistence API. Apress, 2006.

MANFREDI, Silvia Maria. Educação Profissional no Brasil. Cortez, 2003.

MCLAUGHLIN, Brett. Building Java Enterprise Applications. O'Reilly & Associates, 2002.

NICHOLAS, Kassem, Design Entreprise Applications with Java 2 Platform, Entreprise Edition, Addison-Wesley, 2000.

SAMPAIO, Cleuton. Guia do Java Enterprise Edition 5. Brasport, 2007.

WOODS, Dan. Enterprise SOA. O'Reilly & Associates, 2006.

AGRADECIMENTOS

Agradecemos à SETEC/MEC pelo financiamento deste trabalho através do **Projeto de Desenvolvimento**, **Implantação**, **Suporte e Manutenção do Módulo SIEP Gerencial**, **do Sistema de Informações da Educação Profissional e Tecnológica – SIEP**.