



Proposta de um Sistema de Mineração de Dados para Análise do Perfil Socioeconômico e Acadêmico dos Alunos do IFTO – Campus Paraíso do Tocantins

Fábio Silveira Vidal^{1,2}, Rubens dos Santos Silva³

¹Docente do IFTO – Campus Paraíso do Tocantins. e-mail: vidalifs@ifto.edu.br

²Doutorando do Programa de Pós-Graduação em Engenharia de Defesa – Instituto Militar de Engenharia. e-mail: vidalifs@ifto.edu.br

³Discente do Curso de Gestão da Tecnologia da Informação – IFTO/Campus Paraíso. Bolsista do CNPq. e-mail: rubensapollo@hotmail.com

Resumo: O uso adequado da informação, no planejamento estratégico, pode ser um fator determinante para o sucesso de uma instituição. As grandes organizações desempenham grandes esforços para o bom uso de suas bases de dados a fim de fomentar um bom aparato de apoio à decisão. Para tanto, são desenvolvidos e utilizados sofisticados sistemas de mineração de dados. Este trabalho apresenta uma proposta de desenvolvimento de um sistema de mineração de dados para o Instituto Federal do Tocantins – *Campus Paraíso do Tocantins*, com o objetivo de auxiliar os processos de planejamento estratégico da instituição via descoberta de informações e padrões a partir dos históricos escolares e perfis socioeconômicos dos estudantes. Foram feitas entrevistas em setores pontuais, e a partir de informações e documentos adquiridos, se desenvolveu um modelo de banco de dados para a consolidação das informações em uma base de dados única, de forma a viabilizar a correlação de dados via ferramentas específicas para mineração de dados. Tal base de dados fora implementada com o uso do Sistema de Gerenciamento de Banco de Dados MySQL *Community Edition*.

Palavras-chave: Mineração de Dados, Perfil Acadêmico, Perfil Socioeconômico

1. INTRODUÇÃO

O Instituto Federal de Educação, Ciência e Tecnologia do Tocantins (IFTO), integrante da Rede Federal de Ensino Profissional e Tecnológica, é uma instituição de educação básica e superior, pluricurricular, multicampi e descentralizada, ofertando educação profissional e tecnológica nas diferentes modalidades de ensino, com base na conjugação de conhecimentos técnicos e tecnológicos na sua prática pedagógica (INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO TOCANTINS, 2011).

Situado na cidade Paraíso do Tocantins, o *Campus Paraíso do Tocantins* oferece cursos profissionalizantes nas modalidades Integrado ao Ensino Médio e Subsequente, cursos de nível superior e 1 (um) curso de pós-graduação *latto sensu*; além de cursos profissionalizantes de nível médio na modalidade subsequente a distância.

O *Campus* atende alunos oriundos de diferentes realidades social e econômica. Muitos destes alunos tem dificuldades de adaptação ao sistema de ensino do IFTO, o que acarreta em uma concentração da taxa de evasão escolar na etapa inicial dos cursos oferecidos.

A estrutura de Tecnologia da Informação disponível provê acesso à internet e computadores para todos os setores administrativos e de ensino. Entretanto, não há um sistema de informações específico para o gerenciamento acadêmico. As informações a respeito de alunos, professores, funcionários e processos administrativos, ficam organizadas nos setores pertinentes.

A realidade exposta dificulta os processos de tomada de decisão e retardam a emissão de relatórios específicos em cada setor. Segundo JÚNIOR (2006), com o advento da globalização, a competitividade entre as empresas no mundo dos negócios vem aumentando intensamente nos últimos anos. As empresas contemporâneas, cientes da necessidade de adaptação a este cenário, têm investido na captação, armazenamento, tratamento e aplicação da informação como diferencial estratégico e competitivo na condução dos negócios. As organizações que não investem em tal sentido, tendem a evoluir de forma mais lenta, devido à baixa eficácia de análise dos processos pertinentes à sua realidade. Por consequência, perdem oportunidade de corrigir falhas e de implantar projetos e medidas que viabilizem o próprio crescimento.



Entre as grandes preocupações, no âmbito da Rede Profissional de Educação Profissional e Tecnológica, estão a qualidade de ensino, a minimização da evasão escolar, e a maximização do número de alunos. As instituições desta Rede firmaram um acordo de metas com o Ministério da Educação, no qual se deve conseguir um índice mínimo de 90% de permanência até o ano de 2016; e uma relação de 20 (vinte) alunos por professor. Acredita-se que os dois aspectos citados possam ser influenciados por fatores socioeconômicos dos estudantes, como renda familiar, condições de moradia, estrutura familiar, entre outras. Relacionar o baixo rendimento, dificuldades de aprendizagem e evasão escolar com a situação socioeconômica dos estudantes, de modo eficaz, é uma tarefa excessivamente complexa sem o auxílio de um sistema de informação eficiente.

Este trabalho apresenta a proposta de um projeto para o desenvolvimento de um sistema de mineração de dados, reunindo informações registradas em dois dos setores do IFTO – *Campus* Paraíso do Tocantins: (a) COTEPE (Coordenação Técnico Pedagógica); e (b) CORES (Coordenação de Registros Escolares). Tal projeto tem o objetivo de descobrir a existência de fatores socioeconômicos que determinem ou influenciem o rendimento escolar dos discentes; viabilizando tomadas de decisões e/ou ações estratégicas a fim de minimizar os impactos de tais fatores, no sentido de solucionar problemáticas institucionais.

Na próxima sessão, apresenta-se Material e Métodos, abordando entrevistas feitas em setores da instituição, a modelagem de dados e a implementação do banco de dados. Na sessão 3 (três), são mostrados os resultados obtidos com o desenvolvimento do trabalho. Posteriormente, são feitas as conclusões relativas aos resultados e perspectivas de continuidade do trabalho. Por fim, se faz os devidos agradecimentos e referências.

2. MATERIAL E MÉTODOS

2.1 Entrevistas

A primeira etapa do desenvolvimento deste trabalho foi a realização de entrevistas nos setores CORES e COTEPE. Tais setores registram informações a respeito dos discentes da instituição. O primeiro guarda informações a respeito da vida acadêmica dos estudantes, como notas, frequência, matrículas, requerimentos, entre outros. O segundo guarda informações relativas à vida pessoal do aluno como déficits de aprendizagem, renda familiar, número de pessoas na família, doenças, alergias; e o histórico de ocorrências de caráter disciplinar do aluno, normalmente relatadas por professores ou seus próprios colegas.

A CORES faz o controle do desempenho acadêmico dos alunos via planilhas eletrônicas. Em tais planilhas vão informações de aspecto pessoal, como nome, endereço, CPF (Cadastro de Pessoas Físicas), número de identidade, título eleitoral, telefones, *email*, foto, entre outros; e informações de aspecto acadêmico, como número de matrícula, curso em que o aluno está matriculado, situação da matrícula do aluno, notas obtidas nas disciplinas cursadas, frequência em cada disciplina, entre outros. Entretanto, são utilizadas diversas planilhas e arquivos físicos e digitalizados para registro das informações citadas. Dificultando, assim, tarefas rotineiras como consultas e emissão de relatórios.

Similarmente à CORES, a COTEPE também guarda as informações a respeito dos alunos em arquivos físicos e em arquivos em meio digital como planilhas eletrônicas ou documentos de texto. Tarefas rotineiras como consultas e emissão de relatórios também ficam prejudicados no setor.

Ao final das entrevistas, foram coletados modelos de formulários físicos e de arquivos digitais nos dois setores; e percebeu-se redundância entre eles. A partir da análise de tais documentos físicos e digitais, foi feita a modelagem de dados para a implementação de um sistema de banco de dados.

2.2 Modelagem de Dados

O modelo de dados utilizado foi o relacional. De acordo com SILBERSCHATZ et al (2006), este é o principal modelo de dados para aplicações comerciais de processamento de dados. Ele conquistou sua posição de destaque devido à sua simplicidade, que facilita o trabalho do programador, comparado com outros modelos de dados, como o modelo de rede ou o modelo hierárquico.



As primeiras implementações comerciais do modelo relacional se tornaram disponíveis no início da década de 1980, como o sistema SQL/DS (*Structured Query Language/Data System*) no sistema operacional MVS (*Multiple Virtual Storage*), da IBM (*International Business Machines Corporation*), e o SGBD (Sistema de Gerenciamento de Banco de Dados), da Oracle. Desde então, o modelo foi implantado em uma grande quantidade de sistemas comerciais. Os SGBDs relacionais (SGBDRs) populares atuais incluem o DB2 e Informix Dynamic Server (da IBM), o Oracle e Rdb (da Oracle), o Sybase SGBD (da Sybase) e o SQLServer e Access (da Microsoft). Além disso, vários sistemas de código aberto, como MySQL e PostgreSQL, estão disponíveis (ELMASRI et al, 2011).

A fim de organizar o trabalho e facilitar atividades de manutenção e atualização do banco de dados utilizou-se o diagrama de modelo relacional. Em tal diagrama, podem-se visualizar todas as tabelas, seus campos e tipos de dados, índices, chaves e relacionamentos.

Para a construção do Diagrama de Modelo Relacional foi utilizada a ferramenta administrativa MySQL Workbench 5.2, a qual, segundo MYSQL (2012) é uma ferramenta unificada para projetistas de banco de dados, desenvolvedores e administradores de banco de dados. A ferramenta provê suporte para modelagem de dados, desenvolvimento SQL (*Structured Query Language*), ferramentas administrativas para configuração de servidores, administração de usuários, entre outros. O *software* está disponível para os sistemas operacionais Windows, Linux e Mac OS. Trata-se de uma aplicação de uso livre e de código aberto, disponibilizado sob a Licença GPL (*General Public License*).

A GPL é uma licença de *software* elaborada pela GNU *Operating System*. Segundo GNU (2012), Os *softwares* que possuem tal licença podem ser copiados, utilizados e distribuídos sem restrição; além disso, pode-se, inclusive, ter acesso ao código-fonte e alterá-lo para adequações de acordo com necessidades ou preferências.

2.3 Sistema de Gerenciamento de Banco de Dados

O IFTO – *Campus* Paraíso do Tocantins atende cerca de 780 alunos com a oferta de um total de 15 (quinze) cursos. Sendo 8 (oito) profissionalizantes de Nível Médio, 4 (quatro) de Nível Superior, e 1 (um) de Pós-graduação *Lato Sensu*. Além de 2 (dois) cursos profissionalizantes de Nível Médio, na modalidade a distância. A partir de tais números, nota-se que o volume de dados a ser armazenado é de ordem pequena.

O Sistema de Gerenciamento de Banco de Dados (SGBD) escolhido foi o MySQL em conjunto com a ferramenta administrativa MySQL Workbench. Tal SGBD possui diversas edições, sendo a *Community Edition* disponibilizada sob a Licença GPL. Apesar de não dispor de recursos mais avançados como as demais edições do MySQL e outros Bancos de Dados de uso comercial, o MySQL *Community Edition* é gratuito, exige poucos recursos computacionais e possui capacidade suficiente para comportar a demanda exigida para as atividades de mineração de dados a serem realizadas neste trabalho. Sendo, portanto, adequado para tal.

2.4 Mineração de Dados

Metodologias de mineração de dados têm sido desenvolvidas para exploração e análise, por processos automáticos ou semiautomáticos, de grandes quantidades de dados para a descoberta de padrões e regras significativas. Na verdade, tais dados, incluindo dados pessoais, podem fornecer uma rica fonte para descoberta de conhecimento e apoio à decisão. A mineração de dados envolve várias técnicas, incluindo as estatísticas, redes neurais, técnicas de árvore de decisão, algoritmo genético, e visualização da informação que foram desenvolvidas ao longo dos anos (CHUEN et al, 2008).

Além da área educacional, a mineração de dados vem sendo utilizada em outros segmentos, como o *marketing*, detecção de fraudes, medicina, instituições governamentais, bolsa de valores, controle de qualidade, e instituições financeiras. As técnicas citadas podem ser utilizadas para a efetivação de diferentes tipos de trabalhos, como agrupamento, associação, classificação, e predição. Maiores informações sobre os mesmos podem ser encontradas em FAYYAD et al (1996).

3. RESULTADOS E DISCUSSÃO

ISBN 978-85-62830-10-5

VII CONNEPI©2012



A partir da compilação das informações adquiridas em entrevistas nos setores CORES e COTEPE, foi desenvolvido um Diagrama de Modelo Relacional de um sistema de banco de dados para armazenamento das informações contidas em tais setores da instituição. Tal diagrama foi dividido em dois pacotes: (a) Acadêmico, vide Figura 1; e (b) Socioeconômico, vide Figura 2.

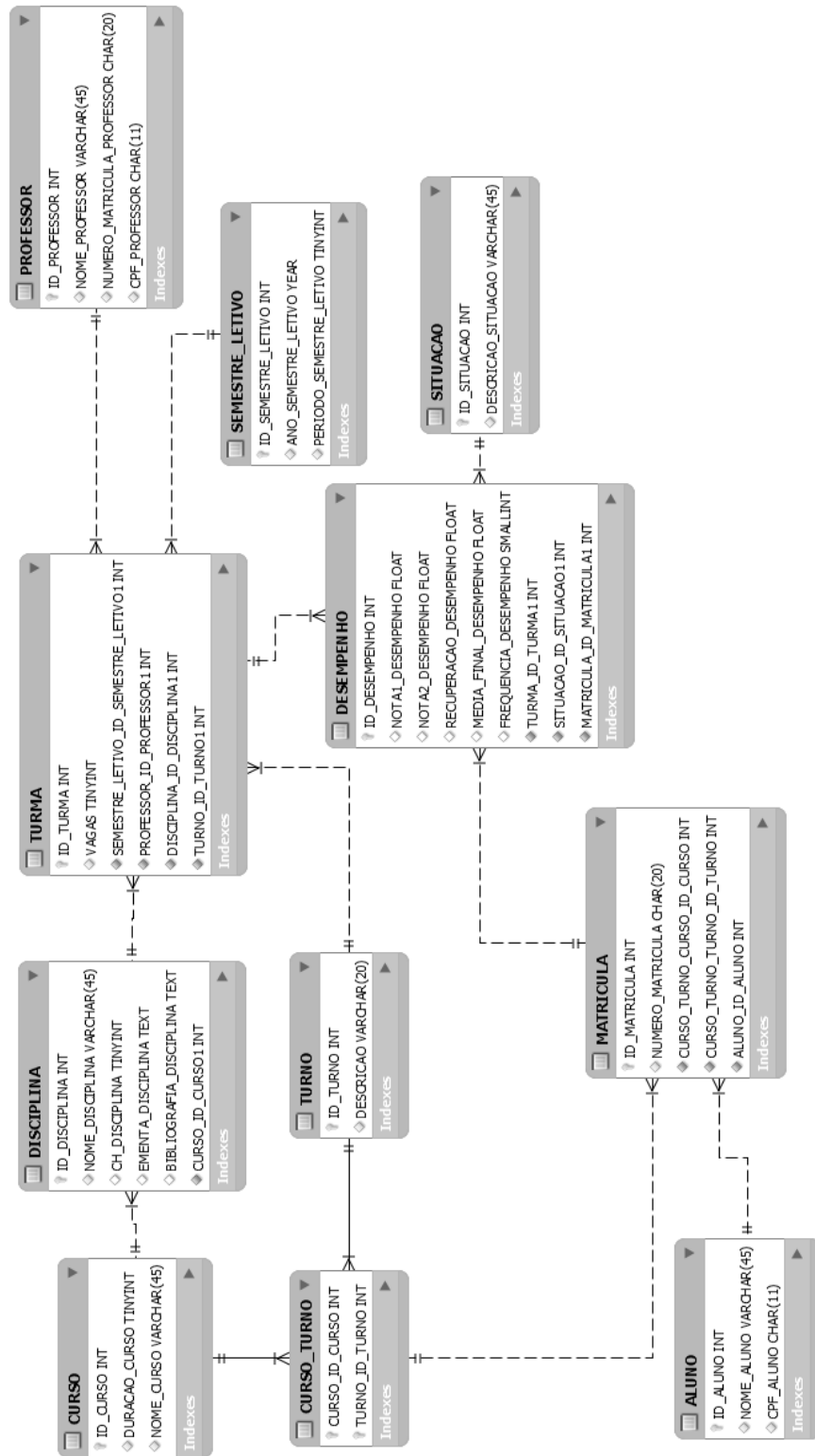


Figura 1- Diagrama de Modelo Relacional - Pacote Acadêmico

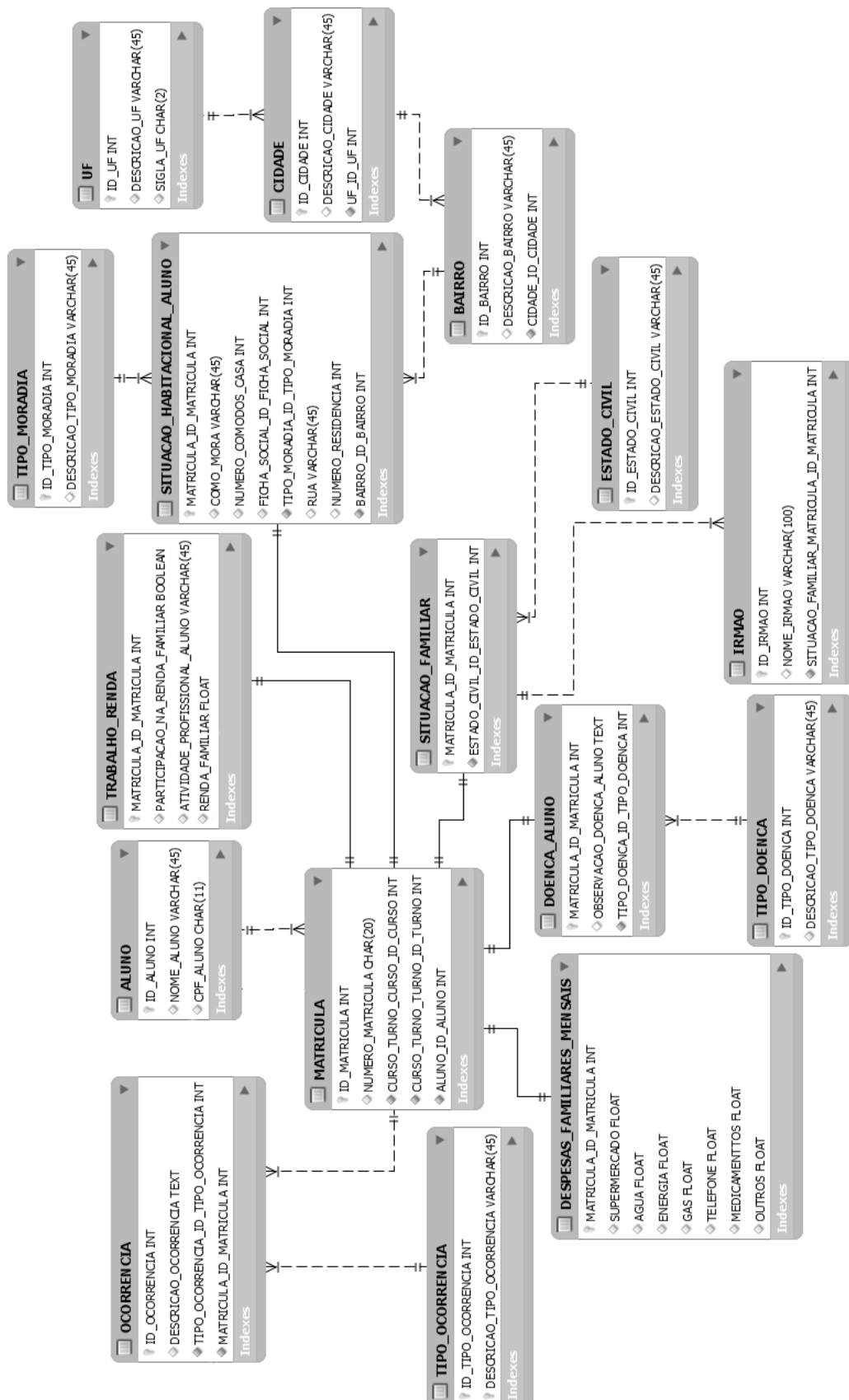


Figura 2 - Diagrama de Modelo Relacional - Pacote Socioeconômico



O Pacote Acadêmico reúne informações a respeito de alunos, matrículas, cursos, disciplinas, turmas, professores, desempenho escolar (notas e frequências dos alunos). Já o Pacote Socioeconômico é voltado para informações a respeito da renda familiar do aluno, sua moradia, situação familiar, tamanho da família, gastos mensais, doenças, e ocorrências (registros feitos na COTEPE com relatos a respeito de desvios de conduta, reclamações ou dificuldades de aprendizagem identificadas a respeito dos alunos).

A fim de otimizar o processamento durante a atividade de mineração de dados, foram criadas duas *views* no banco de dados. A primeira, chamada DM_IFTO_ACADEMICO, para o Pacote Acadêmico, com os campos ID_MATRICULA, ID_CURSO, ID_TURNO, NOTA1, NOTA2, RECUPERACAO, MEDIA_FINAL, FREQUENCIA, ID_SITUACAO, ID_TURMA, e ID_DISCIPLINA; e a segunda, chamada DM_IFTO_SOCIOECONOMICO, para o Pacote Socioeconômico, com os campos ID_MATRICULA, ID_TIPO_OCORRENCIA, ID_ALUNO, PARTICIPACAO_NA_RENDA_FAMILIAR, ID_ESTADO_CIVIL, N_IRMAOS, NUMERO_COMODOS_CASA, ID_TIPO_MORADIA, BAIRRO_ID_BAIRRO, ID_CIDADE, ID_UF, DESPESAS, e RENDA_FAMILIAR. Assim, foram eliminadas informações textuais do conjunto de dados, as quais são de difícil processamento pelos algoritmos de Mineração de Dados. Além disso, também é garantida a confidencialidade dos dados, uma vez que informações sensíveis como o CPF do aluno ou o nome do professor não constam nas *views* apresentadas. O banco de dados deve ser utilizado de forma integral apenas para emissão de relatórios ou para atividades de pós-processamento.

6. CONCLUSÕES

O banco de dados desenvolvido neste trabalho viabiliza a captação de informações estratégicas a respeito da vida escolar e social de uma instituição de ensino. Como estudo de caso, foi utilizado o IFTO – *Campus* Paraíso do Tocantins.

Com o sistema proposto, anseia-se conseguir relacionar as condições dos alunos em sua vida fora da instituição com o seu desempenho e comportamento dentro dela. Após se fazer tal descoberta de informação, facilitar-se-á tomadas de decisões e o planejamento de ações efetivas e estratégicas para melhoramento dos indicadores de eficiência da instituição, como as taxas de aprovação e evasão, e a relação aluno/professor, entre outros.

O sucesso do trabalho de uma instituição de ensino leva a um notório impacto social. Ao passo que mais pessoas consigam melhorar seus níveis de educação e capacitação, o nível de renda da população tende a crescer e movimentar a economia local; oportunizando novas oportunidades de emprego e o crescimento da própria instituição.

Para a continuidade do trabalho, é necessário o preenchimento do banco de dados com informações reais contidas nos setores CORES e COTEPE. Os dois setores utilizam arquivos físicos ou digitais, o que inviabiliza a importação automática de dados. Portanto, será desenvolvido um sistema de entrada de dados.

Com a base de dados formada, outra atividade a ser executada é a utilização de ferramentas específicas para mineração de dados como WEKA, vide HALL et al (2009), ou KELL, vide ALCALÁ-FDEZ et al (2011); o que viabilizará o alcance dos objetivos do projeto, bem como aplicar a mesma metodologia em outras instituições de ensino.

AGRADECIMENTOS

Os autores do trabalho agradecem: (a) ao IFTO, pela oportunidade de realização de trabalho e pelo suporte oferecido para a execução das atividades relacionadas ao projeto; (b) ao CNPq, pelo apoio financeiro em forma de bolsa de estudo; e (c) aos funcionários dos setores CORES e COTEPE pelas informações e modelos de documentos cedidos durante as entrevistas.



REFERÊNCIAS

ALCALÁ-FDEZ, J.; FERNÁNDEZ, A.; LUENGO, J.; DERRAC, J.; GARCIA, S.; SÁNCHEZ, L.; HERRERA, F. KEEL **Data-Mining Software Tool: Data Set Repository, Integration of Algorithms and Experimental Analysis Framework**. In *Journal of Multi-Valued Logic & Soft Computing*, Volume 17, 2011.

CHUEN, C.; CHEN, L. **Data Mining to Improve Personnel Selection and Enhance Human Capital: A Case Study in High-technology Industry**. In: *Expert System with Applications*, Volume 34, 2008.

ELMASRI, R.; NAVATHE, S.B. **Sistemas de Banco de Dados**. 6ª Edição. São Paulo: Pearson Addison Wesley, 2011.

GNU Operating System. **GNU General Public License v2.0 - GNU Project - Free Software Foundation**. Disponível em: <http://www.gnu.org/licenses/old-licenses/gpl-2.0.html>. Acesso em 07 de ago 2012.

HALL, M.; FRANK, E.; HOLMES, G.; PFAHRINGER, B.; REUTEMANN, P.; WITTEN, H. I. **The WEKA Data Mining Software: An Update**. In: *ACM SIGKDD Explorations Newsletter*, Volume 11, 2009.

INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO TOCANTINS. **Relatório de Autoavaliação - Exercício 2010**. Palmas, 2011.

JUNIOR, F.G.; RIBEIRO, L.S.; ROQUE, U.; JOB, D.H.; VIEIRA, E.S.; GOLDSCHMIDT, R.R. **Datawarehouse e Mineração de Dados nos Institutos Superiores da FAETEC**. In: *II Workshop De Computação Científica da UENF*, 2006.

MYSQL. **MySQL :: MySQL Workbench 5.2**. Disponível em: <http://www.mysql.com/products/workbench/>. Acesso em 07 de ago 2012.

SILBERSCHATZ, A.; KORTH, H.F.; SUDARSHAN, S. **Sistemas de Banco de Dados**. 5ª Edição. Rio de Janeiro: Elsevier, 2006.