**RELATÓRIO DE ATIVIDADES DESENVOLVIDAS – DOUTORAMENTO – UMINHO**

**NO TRIMESTRE MAIO-JUNHO-JULHO**

**DOUTORANDO: JOSÉ ANTONIO DA CUNHA**

* **Coleta de dados do sistema acadêmico**

Após análise da base de dados do sistema acadêmico, ficou evidente que seria mais viável extrair os dados, através dos QAcademico, uma vez que o DER do banco de dados, está deveras normalizado, contendo mais de 350 tabelas, muitas das quais, tendo apenas, dois atributos, a chave primária e outro campo de dados. Tornando impraticável extrair os dados dos mesmos.

Por outro lado, o sistema QAcademico, me permite extrair diversos relatórios em formatos de planilhas ou arquivos XML, basta apenas que eu tenha senha com privilégios adequados. Após conversa com o responsável pelo sistema, o mesmo me forneceu uma senha com tais privilégios.

A partir de então, eu pude extrair os seguintes dados do sistema acadêmico:

* Dados dos alunos
* Dados da matriz-curricular
* Dados do boletim-escolar
* Dados dos cursos habilitados
* Dados dos Históricos por cursos
* Dados dos Professores
* Dados das Disciplinas
* Índice de aluno x professor
* Dados dos percentuais de conclusão de cursos
* Dados sócios-econômicos
* Dados de acesso a biblioteca
* Dados de acesso ao posto médico

Os dados acima, foram coletados no formato de planilha Excel e, apenas foram coletados os dados de 2010, 2011, 2012 e 2013, do campus NATAL Central.

Com os dados nas planilhas, então, passa-se para a segunda fase, que é a carga desses dados para o banco de dados.

* **ETL dos dados para o modelo multidimensional**

Acho desnecessário criar um modelo relacional para receber a carga desses dados e, depois criar um modelo dimensional e, transferir esses dados do modelo relacional para o multidimensional. Então criei diretamente um modelo multidimensional e fiz o ETL direto das planilhas para o mesmo. A Figura 1 mostra o modelo multidimensional parcial do sistema. Para realizar o ETL, desenvolvi pacotes do analisys Services, para que o processo fique dinâmico, ou seja, feito o pacote e salvo no sistema, pode-se executá-lo, quantas vezes for necessário.



**Figura 1 Modelo Multidimensional Parcial do Sistema SISEDUCACIONAL**

* **Criar as consultas SQL para popular a Tabela de fato (FatoMatriculaAluno)**
* Trabalhei orientando três alunos na produção de três artigos, para serem submetidos, possivelmente a revista Holos ou, talvez em algum congresso da área. Os artigos são:
* ALGORITMOS GENÉTICOS PARA O DESENVOLVMENTO DE UM SISTEMA DE DISTRIBUIÇÃO E GERENCIAMENTO DE TAREFAS PARA METODOLOGIA SCRUM
* Avaliação do aprendizado em ambientes Educação à distância USANDO MINERAÇÃO DE DADOS
* PROPOSTA ARQUITETURAL DE SISTEMA INTELIGENTE DE ACONSELHAMENTO PEDAGÓGICO APLICANDO RACIOCINIO BASEADO EM CASOS
* Foi desenvolvido um protótipo preliminar do sistema (Tela principal). A Figura 2 mostra a tela principal, versão 0 do sistema.



**Figura 2 Tela principal do sistema.**

Muitas dessas atividades estão sendo desenvolvidas em paralelo, como por exemplo, a modelagem do sistema, o ETL e RBC. O Data Warehouse e o Data Mining, dependem da fase de ETL, só após esta fase concluída é que pode-se partir para essas duas etapas.

* **O módulo de Raciocínio Baseado em Casos.**

Este módulo está sendo desenvolvido, onde já foi definido o modelo de formulário do modelo de Caso a ser utilizado, bem como os algoritmos de indexação e recuperação dos Casos na base de Casos (algoritmo ***K-means clustering***). Também já está definida a tecnologia que será usada para ser a Base de Casos (Base de Dados ***NoSQL***), pois facilita o manuseio com objetos ***Json*** direto no código.

**PROPOSTA ARQUITETURAL DE SISTEMA**

* **Linguagem de Programação**

Propoe-se utilizar uma linguagem de programação de ampla aceitação, que permita a implementação independentemente da plataforma utilizada. Dentre as mais conhecidas, a linguagem **Javascript**, na sua especificação mais recente, é linguagem de programação poderosa, suficientemente complexa e que permite sua aplicação tanto compilada em servidores e máquinas virtuais, quanto como aplicações clientes, sendo interpretadas em navegadores de internet comuns.

Dentre as opções possíveis, o Javascript é a linguagem candidata perfeita, visto que pode ser aplicada em navegadores de internet, em sistemas operacionais de computadores, smartphones e tablets.

* **Persistência de Dados**

Como um fator importante ao RBC, a base de casos sendo persistida em um banco de dados que mantém sua estrutura, como coleção de documentos, e ainda permitir a utilização nativa de algoritmos de classificação, mapeamento e redução, nos auxilia a obter resultados que anos antes, com uso de outras tecnologias seriam mais difíceis e mais custosas de se obter.

Será Proposto a utilização de uma base de dados **noSQL** orientada a documentos como mecanismo de persistência dos casos utilizados na solução proposta. O termo **noSQL** é usado amplamente para descrever estruturas de bancos de dados que vão além da estrutura relacional, e que podem incluir conceitos de tuplas, grafos, pares de chave-valor, árvore, documento etc.

Dentre as soluções disponíveis no mercado, a solução que mais se assemelha ao proposto é um **SGBD** (Sistema de Gerenciamento de Banco de Dados) não-relacional, orientado a documentos, de código aberto e uso livre. O **SGBD MongoDB** fornece nativamente métodos de mapeamento e redução, que permite obter resultados otimizados e maior velocidade de processamento com redução de recursos.

O MongoDB utiliza também o Javascript como linguagem nativa para suas rotinas e aplicações, além de armazenar os dados em um formato semelhante à notação de objetos Javascript (JSON) só que em binário ao invés de texto pleno. A Figura 2 ilustra a comparação entre um documento representado em texto pleno e sua representação em binário utilizando a notação Javascript.



* **Classificação**

Como metodologia de facilitação para indexação e recuperação dos casos, será adotado a classificação dos casos dividindo-os em agrupamentos chamados classes, levando em consideração às demandas relacionadas a cada caso e determinando um peso a cada demanda. Com isso pode-se determinar um valor para cada classe, de acordo com a quantidade de demandas e o valor individual das demandas de cada classe.

Seguindo o conceito dos quatro elementos de um caso ideal: a descrição dos aspectos relevantes do caso; o contexto no qual o caso está inserido; a descrição da solução associada ao caso e a avaliação da solução empregada, definimos os casos dentro do domínio de aconselhamento educacional conforme os termos empregados pelas especialistas consultoras deste trabalho.

Dessa forma, o contexto pode ser considerado como uma ou mais classes de demandas, onde cada demanda descreve um aspecto do caso analisado. As soluções propostas serão compostas por um ou mais encaminhamentos, enquanto que a classificação de cada caso é

composta pelo peso da classe do caso multiplicado pela soma de pesos de cada demanda que compõe o caso.

A classificação em classes de demandas foi realizada manualmente, por um especialista da área do domínio da aplicação. As demandas iniciais estão listadas em ordem alfabética na Lista 1 como segue:

•Atrasos constantes

• Bulling

•Conflito com relação a opção sexual

•Conflito familiar

•Conflito relacionado a situação afetiva, sexual ou de relacionamento

•Desequilíbrio ou caso de ordem psicológica

•Desmotivação pela opção de curso

•Desmotivação por baixo rendimento

•Dificuldade de aprendizagem em disciplina

•Muitas faltas

•Necessidade de orientação pedagógica

•Necessidade de orientação secular (pessoal, não relacionado com a vida acadêmica)

•Pais em separação

•Caso de comportamento

•Caso de ordem disciplinar grave

•Caso de ordem disciplinar leve

•Caso de ordem disciplinar médio

•Caso de relacionamento aluno x aluno

•Caso de relacionamento com a mãe

•Caso de relacionamento com o pai

•Caso de relacionamento em casa

•Caso de relacionamento professor x aluno

•Caso de ordem socioeconômica do aluno ou da família

•Situação de abuso (moral, sexual etc.)

•Situação de exclusão em sala

•Situação relacionada à timidez

As classes resultantes desta classificação são distintas na natureza de suas demandas e também na quantidade de demandas que compõe cada classe. Também foi atribuído um valor (peso) para cada demanda, baseado em seu nível de necessidade de atenção, levando ao prognóstico de que demandas com valores mais altos dentro de cada classe carecem de atenção maior por parte dos especialistas do que as que tem valor menor.

Da mesma forma, as classes de valores mais altos necessitam de maior atenção. Com este mecanismo de classificação simples, conseguimos distinguir casos menos graves, aqueles com demandas com valores baixos pertencentes a classes com valores menores dos casos mais graves (que possuem demandas e pertencem às classes com valores mais altos).

* **O módulo do Aluno**

Este módulo está sendo desenvolvido um *game compettition* por um aluno sobre minha orientação. No entanto, ainda não foi testada a integração com o sistema ASP.NET MVC. A tecnologia que está sendo utilizada é o ***Construct 2***. Neste game compettition o Aluno seleciona o assunto que o mesmo está interessado em melhorar seu desempenho escolar e, é orientado a desenvolver certas atividades competitivas. Ele pode jogar sozinho ou em grupo de no máximo 4 pessoas.

**Dúvidas**?

* Tiro ou deixo o tópico sobre Redes Neurais?
* Acrescento o tópico sobre Banco de Dados NoSQL?