

## **ANÁLISE DAS TECNOLOGIAS DE BANCO DE DADOS PARA INTEGRAÇÃO DE DADOS ESTRUTURADOS E SEMI-ESTRUTURADOS**

Ingrid Cavalcanti XAVIER<sup>1</sup>; José Laurindo CAMPOS DOS SANTOS<sup>2</sup>; Sidinéia Aparecida AMADIO<sup>3</sup>.

<sup>1</sup>Bolsista PIBIC/CNPq/INPA; <sup>2</sup>Orientador GTI- COAE /INPA; <sup>3</sup>Co-orientadora CPBA/ INPA.

### **1. Introdução**

As extensas áreas e a abundância de recursos naturais na Amazônia permitem que a região tenha uma ictiofauna (peixes) diversificada, distribuída por centenas de igarapés, lagos e rios. Para preservar as informações dessa imensa riqueza, pesquisadores do Instituto Nacional de Pesquisas da Amazônia (INPA) realizam expedições de coletas de peixes em lugares específicos como Balbina, Catalão, entre outros. Durante a análise laboratorial, no INPA, são coletados os dados dos peixes, verificando os detalhes de cada espécie e após isso, são inseridos em recursos de planilhas eletrônicas onde os dados coletados são de estrutura heterogênea. Se mantidos nessas estruturas será de difícil análise e acesso aos resultados de pesquisa.

Com o entendimento deste problema, surgiu a necessidade de prover mecanismos que, possam facilitar um gerenciamento integrado dos dados biológicos, tanto no contexto dos experimentos quanto no gerenciamento de coleções. Criou-se então o projeto denominado "Desenvolvimento de uma ferramenta de integração de banco de dados/ metadados bio-ecológicos das comunidades de peixes da Amazônia Central" que tem como objetivo desenvolver uma ferramenta que solucione os problemas de análise e manipulação dos dados em bases heterogêneas, onde se encontram armazenados os dados biológicos referentes a coletas de peixes da Amazônia Central, de forma que os dados possam ser visualizados unificadamente.

### **2. Material e métodos**

O procedimento metodológico e a técnica adotados inicialmente foram pesquisas e revisões bibliográficas com a finalidade de um melhor conhecimento de aspectos relacionados ao projeto, como por exemplo, o estudo de dados/metadados e a verificação de um SGBD (Sistema Gerenciador de Banco de dados) compatível com o projeto, o qual foi selecionado dois para análise: PostgreSQL e o MySQL. O estudo do ambiente bio-ecológico e o levantamento de requisitos foram métodos imprescindíveis ao projeto. O primeiro se deu com a verificação do ambiente de trabalho dos pesquisadores no momento em que os peixes chegam ao laboratório até a inserção dos dados no banco de dados atual (Access), além da verificação posterior desse ambiente de gerenciamento. O segundo método foi realizado mediante entrevistas com a Dra. Sidinéia Amadio, uma das pesquisadoras que gerenciam os dados bio-ecológicos. Por meio dessas etapas, foram identificados os problemas, encontradas novas funcionalidades e criado o Esquema Conceitual (Diagrama Entidade e Relacionamento) – *Figura 1*. Após o procedimento dessas atividades, foram criados os Esquemas Lógico e Físico e a implementação dos dados no SGBD PostgreSQL 8.3. Tendo em vista a implementação de uma ferramenta de integração, foram realizados estudos de CSS (*Cascading Style Sheets*) para definir a apresentação de documentos escritos na linguagem de marcação HTML (*HyperText Markup Language*) e pesquisas sobre JSP (*JavaServer Pages*) que é uma linguagem de programação muito utilizada para gerar conteúdo dinâmico na World Wide Web (Oliveira, 2004).

### **3. Resultados e discussão**

O INPA possui projetos científicos de Coleções de peixes, que tem como objetivo geral identificar os padrões de variação da biologia do peixe e parâmetros biológicos em relação às mudanças ambientais (variações interanuais). Para tanto, pesquisadores realizam expedições de coletas de peixes em lugares específicos como Balbina, Catalão, entre outros. Em uma excursão, um ou mais pesquisadores vão a campo realizar a pesca, coletando diversos peixes de diferentes espécies. Eles usam uma ficha padronizada para campo, anotando alguns dados observados. Os coletores podem ser os pesquisadores ou pescadores. A metodologia de campo é, além da coleta de peixes, verificar dados como: o local de coleta, data, hora, posição geográfica do local e aparelhos utilizados na pesca, como malhadeira, redinha, tarrafa, entre outros. A tarrafa, por exemplo, é um material apropriado para águas rasas, águas turvas e corredeiras. A malhadeira é um aparelho padronizado pelos pesquisadores do INPA, sendo inicialmente usada uma bateria para coleta, correspondente a 10 malhadeiras. A partir do mês de abril de 2007, surgiu outra classificação chamada de aparelho extra, ou 2 baterias, correspondentes a 20 malhadeiras. É importante identificar quais são os

peixes coletados da malhadeira inicial (1 bateria), e os da malhadeira extra, para uma melhor análise. Após a coleta, os peixes são levados para estudo no INPA, onde são verificadas suas características. Essa fase é conhecida como pós-coleta. No projeto, são coletadas informações sobre o peixe e registrado se é testemunha ou não. Se for, é depositado na coleção. Todo trabalho biológico, ecológico, entre outros, exige a "testemunha", que é de um a três exemplares de cada espécie que são depositados na coleção zoológica. Uma coleção só pode existir em locais oficiais, e o INPA é um fiel depositário, pode receber peixes testemunhos de toda parte do Brasil. A distribuição dos exemplares (peixes) pode ser realizada de diferentes modos, sendo classificados como testemunhas e depositados na coleção, ou descartados do projeto. Para as pesquisas realizadas, é importante armazenar dados como o ano hidrológico, lembrando que este é diferente do ano gregoriano, pois não começa em janeiro e termina em dezembro, isso depende do período hidrológico, ou seja, para completar um ano hidrológico é preciso passar por quatro períodos: enchente, seca, vazante e cheia, os quais não têm um tempo de duração pré-definido. Quanto à localização que o peixe se encontra é utilizado o GPS, ou seja, são verificadas as coordenadas. No caso do projeto Catalão, o local está atualmente padronizado em um ponto GPS no Lago do Padre, porém, para outros projetos eles utilizam mais de um ponto GPS em um mesmo local. A morfometria do peixe, ou seja, as medidas do peixe, são inicialmente o comprimento padrão e total. A taxonomia também é verificada, ou seja, ordem, gênero, espécie, família. Podem ser registradas multimídias dos indivíduos, por exemplo, fotos e vídeos. Em laboratório são arquivadas caracterizações como gordura, peso total do peixe, peso eviscerado, peso do fígado, estômago, peso da gônada, peso da amostra da gônada. Além de serem verificadas características dos peixes como um todo, é feito um posterior estudo de partes específicas, como: escama, estômago e cabeça. Com a escama, faz-se o estudo do crescimento da espécie, com o estômago analisa-se a dieta do peixe e, com a cabeça obtém-se a caracterização da espécie. Para isso, é necessário armazenar quais foram os itens coletados. Além desses estão: DNA, gônada, histologia, olho e otólito. Para uma melhor análise, é importante identificar aspectos abióticos como: temperatura da água, o tipo e pH da água, a temperatura do ar, oxigênio dissolvido, condutividade, pressão da água e o tempo. Os dados dos peixes são correlacionados com o nível da água que ocorre todos os anos, sendo este dado obtido na Portobrás, (porto de Manaus). Essa informação não é um dado armazenado pelos pesquisadores, mas é de muita importância, pois é através dele, que se verifica quando se trata de uma enchente, vazante, seca, ou cheia (referente ao período hidrológico). Como mencionado anteriormente, durante a análise laboratorial são coletados os dados dos peixes verificando os detalhes de cada espécie. Após isso, o que se coletou é inserido em um banco de dados que não possui recursos necessários. A estrutura atual de armazenamento ocasiona muitas desvantagens de análise e gerenciamento, algumas delas são as constantes repetições de dados, como pode ser observado na [Tabela 1](#), pois além de gerar duplicação de esforços, o espaço de armazenamento é desperdiçado. Os dados biológicos referentes a coletas de peixes da Amazônia Central são armazenados em bases heterogêneas, ou seja, são criadas várias planilhas para seu gerenciamento. Observando esses fatos, a solução encontrada utilizando o SGBD contribuirá com esses e outros problemas.

Tabela 1 – Ambiente de gerenciamento Access com alguns dados bio-ecológicos

<b>Hora</b>	<b>Ordem</b>	<b>Família</b>	<b>Gênero</b>	<b>Espécie</b>	<b>Comp (mm)</b>	<b>Peso (gr)</b>	<b>Sexo</b>	<b>Est. Mat</b>
CV	Characiformes	Hemiodontidae	Hemiodus	"rabo de fogo"	180	64,60	M	2
CV	Characiformes	Hemiodontidae	Hemiodus	"rabo de fogo"	150	60,80	F	2
CV	Characiformes	Hemiodontidae	Hemiodus	"rabo de fogo"	171	55,60	F	2
CV	Characiformes	Hemiodontidae	Hemiodus	"rabo de fogo"	179	67,60	F	2
CV	Characiformes	Hemiodontidae	Hemiodus	"rabo de fogo"	165	67,60	F	2

Com a análise dos problemas encontrados foi criado o esquema conceitual (Figura 1). Para melhor visualização o esquema conceitual foi dividido (Figura 1.a). Para que se possam retratar os fatos descritos, e que os mesmos possam nos levar à decisões e ações, se faz necessário então registrá-los. Este registro é feito através da criação de um modelo. O objetivo da Modelagem de dados é representar de forma única, não redundante e resumido, os dados de uma aplicação. Em projetos Conceituais de aplicações em banco de dados o modelo Entidade-Relacionamento é o mais largamente utilizado para a representação e entendimento dos dados que compõem a essência de um sistema (Machado & Abreu, 2006). Avaliando a ferramenta Netbeans IDE 6.5 para a implementação das páginas web, verificou-se vantagens como ambiente de desenvolvimento robusto, facilidade de uso devido suas interfaces. No requisito de facilidade de manipulação, a ferramentas é intuitiva, mostrando os elementos gráficos, por exemplo, botão, quadros e caixa, as propriedades dos elementos criados e o código fonte, dentre outras informações de apoio ao desenvolvimento. Possui desvantagens como o código gerado pelo Visual JSF (*Java Server Faces*) ser um pouco poluído. Também se torna um pouco lento devido a grande quantidade de recursos que possui.

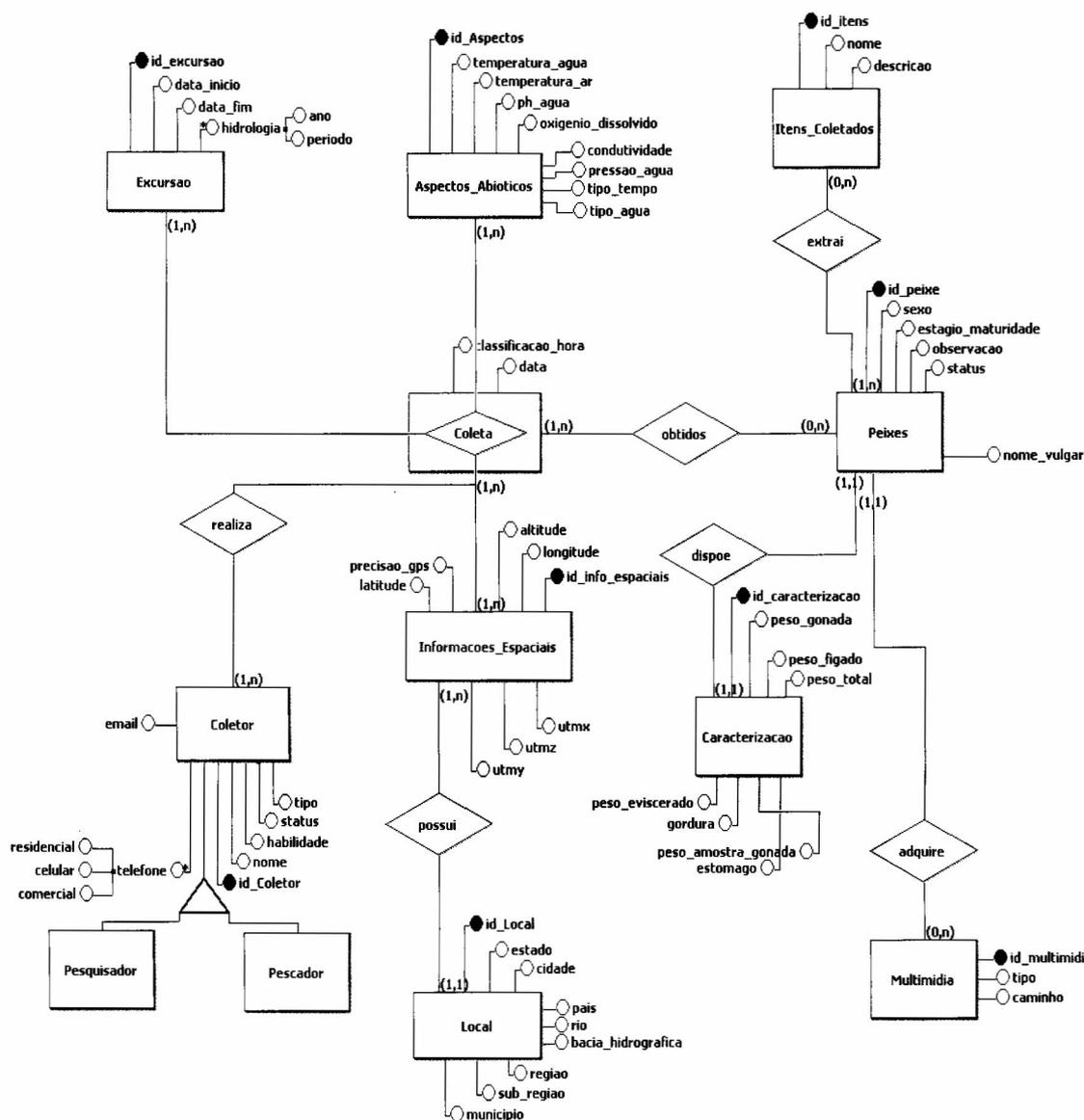


Figura 1. Esquema Conceitual (Diagrama Entidade-Relacionamento)

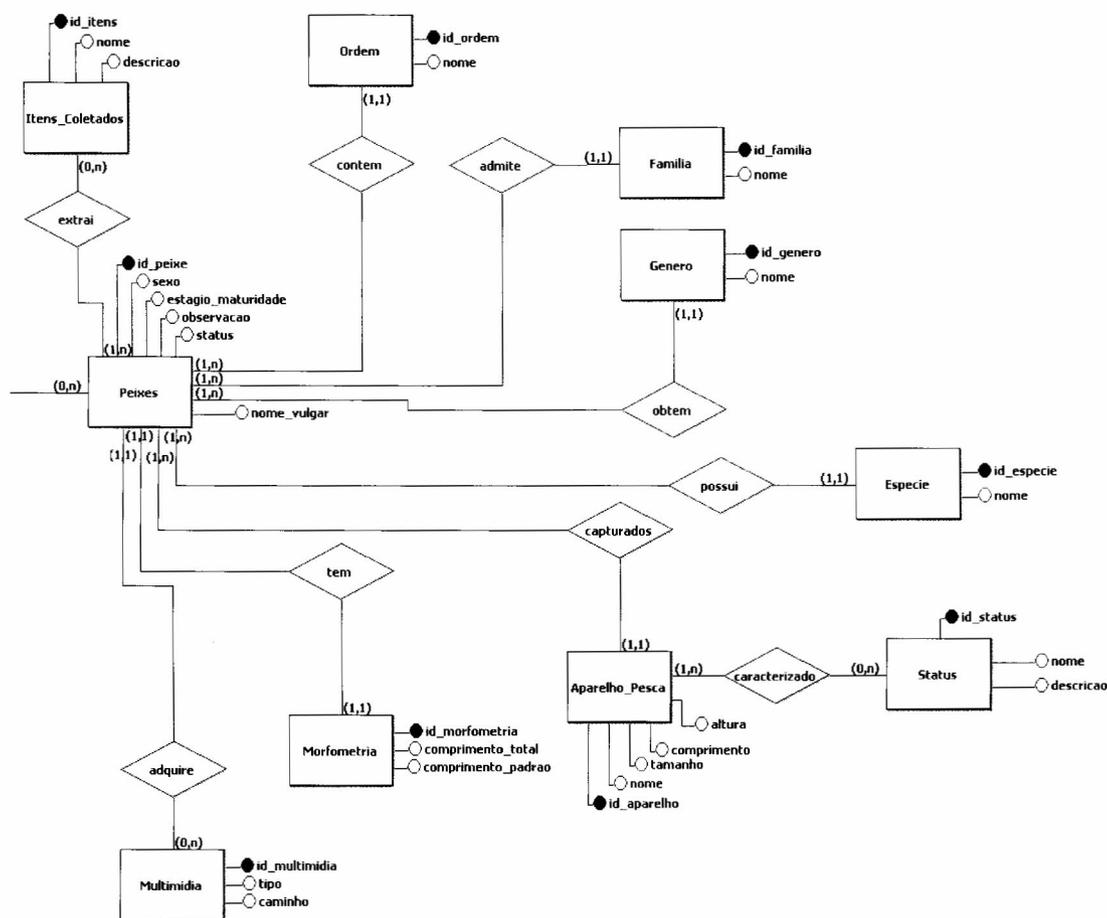


Figura 1.a Esquema Conceitual.

#### 4. Conclusão

O desenvolvimento da ferramenta de integração de banco de dados/metadados tem como objetivo contribuir com soluções para os problemas relacionados à análise, gerenciamento e disseminação da informação. A experiência de desenvolver um trabalho de pesquisa é bastante válida no âmbito profissional e crescimento pessoal, pois se obtém novos conhecimentos por meio de temas não abordados na vida acadêmica, aprendizado no desenvolvimento de um relatório e estímulo para pesquisas científicas. Dificuldades comuns são encontradas no processo, como o entendimento de novos assuntos, o que são resolvidos mediante orientações e estudos/pesquisas em livros e/ou artigos. Também com as regras de negócios do projeto bio-ecológico, resolvido com levantamento de requisitos. Fez-se estudo do ambiente NetBeans IDE 6.5 para desenvolvimento da ferramenta web e foi realizado o relatório científico paralelamente ao desenvolvimento dessas atividades. Espera-se que o sistema web Coleções de Peixes, possa suprir necessidades de usuários finais e servir de referência para outros alunos que estejam em busca de conhecimentos similares. No entanto, como futuras propostas, deseja-se a continuação do trabalho com relatórios e consultas. A importância em desenvolver relatórios está em ter um controle maior sobre os dados, obtendo um nível maior de informação para posteriores análises. Deseja-se também a criação de um terceiro tipo de usuário para realizar apenas inclusão e pesquisa de dados, pois resultaria maior segurança da informação, e agilidade do trabalho por outros funcionários que não possuem senha do administrador, quando estes se encontrarem atarefados.

### **5. Referências**

Oliveira, E. 2004 (setembro). Java Server Pages e Servidor Tomcat. (<http://www.linhadecodigo.com.br/ArtigoImpressao.aspx?id=452>). Acesso: 10/03/2009.

Machado, F & Abreu, M. 2006. *Projeto de Banco de Dados - Uma visão prática*. Editora Érica.