



INSTITUTO TECNOLÓGICO DE AERONÁUTICA  
CE-240 PROJETO DE SISTEMAS DE BANCO DE DADOS  
Prof. Dr. Adilson Marques da Cunha

**Lista de Exercícios 2**

Antônio Magno Lima Espescht  
[antonio.espescht@gmail.com](mailto:antonio.espescht@gmail.com)

São José dos Campos - SP

30 de março de 2009

# Sumário

|          |  |           |
|----------|--|-----------|
| <b>1</b> | <b>Introdução .....</b>  | <b>1</b>  |
| 1.1      | Título .....   | 1         |
| 1.2      | Motivação .....  | 1         |
| 1.3      | Objetivo .....   | 1         |
| <b>2</b> | <b>Conteúdo.....</b>   | <b>2</b>  |
| 2.1      | Heurística do Objetivo Aplicada ao Módulo ADF .....  | 2         |
| 2.1.1    | Tematização .....  | 2         |
| 2.1.2    | Motivação .....  | 2         |
| 2.1.3    | Contextualização .....   | 2         |
| 2.1.4    | Objetivação .....  | 3         |
| 2.1.5    | Intitulação a Partir de Uma Redução de Escopo .....  | 6         |
| 2.1.6    | Especificação de Requisitos.....   | 6         |
| 2.2      | Heurística do Objetivo Aplicada à Proposta de Dissertação de Mestrado .....  | 6         |
| 2.2.1    | Tematização .....  | 6         |
| 2.2.2    | Motivação .....  | 6         |
| 2.2.3    | Contextualização .....   | 7         |
| 2.2.4    | Objetivação .....  | 7         |
| 2.2.5    | Intitulação a Partir de Uma Redução de Escopo .....  | 10        |
| 2.2.6    | Especificação de Requisitos.....   | 10        |
| 2.3      | Pesquisas na Internet .....  | 11        |
| 2.3.1    | Pesquisa sobre Filtros e Cálculos Estatísticos para Séries Hidrológicas<br>(Temática de Protótipo de Projeto de Aplicativo de Banco de Dados)..... | 11        |
| 2.3.2    | Pesquisa sobre Mecanismos de Busca para Parâmetros de Ensaio (Temática<br>de Mestrado) .....   | 11        |
| 2.3.3    | Pesquisa sobre Genealogia (Hobby).....   | 12        |
| <b>3</b> | <b>Conclusão e Comentários.....</b>  | <b>12</b> |
| <b>4</b> | <b>Referências .....</b>   | <b>13</b> |

# 1 Introdução

Em atendimento à Lista de Exercícios (ListEx) 2, foram realizadas as atividades:

- Aplicar a **Heurística do Objetivo** [1] ao módulo **ADF** (Aplicação de Filtros) do *Sistema de Aquisição, Tratamento, Monitoramento e Difusão de Dados Hidrológicos (Sis ATMH)*.
- Aplicar a **Heurística do Objetivo** [1] à Proposta de Pesquisa em Nível de Mestrado “**A Web Search Engine to Flight Test Data**” [2].
- Pesquisar três diferentes assuntos na Internet:
  - O módulo ADF: aplicação de filtros em séries de dados hidrológicos;
  - Minha proposta de mestrado: um mecanismo de busca para dados de ensaios em voo.
  - Meu hobby: Genealogia.

Este texto sumaria os resultados obtidos.

## 1.1 Título

*Módulo de Aplicação de Filtros e Cálculo de Estatísticas (AFE).*

## 1.2 Motivação

Agregar o interesse da *Agência Nacional de Águas (ANA)* em ter um sistema para armazenamento e processamento de séries históricas de dados hidrológicos com a necessidade dos alunos do curso CE-240, do Primeiro Período de 2009, em aprender como projetar sistemas de bancos de dados.

O conhecimento teórico adquirido durante o curso será empregado pelos alunos do curso CE-240, do Primeiro Período de 2009, para que estes implementem o Sis ATMH, até 22/06/2009, implantando este sistema no laboratório da *Fundação Casimiro Montenegro Filho (FCMF)*, neste mesmo prazo.

O Sis ATMH é um Protótipo Acadêmico de Projeto de Aplicativo de Banco de Dados que atende algumas das necessidades da ANA.

## 1.3 Objetivo

Disponibilizar, em tempo hábil para que o sistema completo seja concluído, pelos alunos do curso CE-240, do Primeiro Período de 2009, no laboratório da FCMF e até 22/06/2009, um módulo de software integrável ao sistema Sis ATMH que possua ferramentas de análise visando propiciar o armazenamento de séries históricas, a análise ágil e confiável dos dados armazenados além de permitir a aplicação de filtros, a identificação de *outliers*, o cálculo de estatísticas e o tratamento padronizado de dados espúrios.

## 2 Conteúdo

### 2.1 Heurística do Objetivo Aplicada ao Módulo ADF

Aplicou-se a Heurística do Objetivo no módulo **ADF** (Aplicação de Filtros) do Sistema de Aquisição, Tratamento, Monitoramento e Difusão de Dados Hidrológicos - **Sis ATMH**, obtendo-se os resultados descritos nesta seção.

Considerou-se que, em 23/03/2009 e para compatibilizar o número de módulos a desenvolver com a quantidade disponível de alunos desenvolvedores, as funcionalidade do antigo módulo **CCL** (Cálculo de Limites) do **Sis ATMH** foram incorporadas ao módulo **ADF**.

#### 2.1.1 Tematização

Armazenamento e processamento de séries históricas de dados hidrológicos da Agência Nacional de Águas (ANA).

#### 2.1.2 Motivação

Agregar o interesse da *Agência Nacional de Águas (ANA)* em ter um sistema para armazenamento e processamento de séries históricas de dados hidrológicos com a necessidade dos alunos do curso CE-240 em aprender como projetar sistemas de bancos de dados.

Os conhecimentos teóricos adquiridos em sala de aula serão usados para implementar, em grupo e antes de 22/06/2009, o *Sistema de Aquisição, Tratamento, Monitoramento e Difusão de Dados Hidrológicos (Sis ATMH)* que atende algumas das necessidades da ANA .

#### 2.1.3 Contextualização

Para caracterizar a influência de um rio em uma determinada região é necessário coletar e armazenar diariamente dados hidrológicos durante pelo menos 30 anos.

Estes dados hidrológicos são coletados na forma de extensas séries históricas as quais estão sujeitas a diversos tipos de falhas como, por exemplo:

- Ausência de dados numa determinada data devido à falha do equipamento medidor;
- Erro do funcionário que arbitrou um valor qualquer ao invés de efetivamente realizar a medida;
- Erro de transcrição para a base de dados hidrológicos do dado informado pelo funcionário que realizou a medição;

Alguns erros podem ser identificados com o cálculo de estatísticas. Por exemplo: numa série histórica, podemos calcular a média e o desvio padrão de todos os dados. Em seguida, podemos suspeitar dos valores maiores que a média mais duas vezes o desvio padrão, descartando-os em análises posteriores.

Cálculos estatísticos como mínimo, média, máximo e desvio-padrão também podem ser usados para caracterizar séries históricas permitindo substituir uma extensa lista de dados pelos quatro valores estatísticos que os caracterizam.

Além dos dados inexistentes e possivelmente inexatos há também os *outliers*[8] que são valores corretos mas que se afastam do que poderia ser considerado o comportamento

“normal” dos dados coletados. É o que ocorre, por exemplo, quando um raro terremoto provoca uma grande e súbita redução do fluxo de água em um rio num determinado dia. Esta variação deve ser removida quando se estuda o comportamento “médio” deste rio em um determinado período.

Para reduzir o efeito destes dados errados ou inexistentes, e para amenizar variações bruscas devidos a diversos fatores, costuma-se usar dados “filtrados” ao invés dos dados originais[5]. Por exemplo, quando se aplica um filtro tipo “média móvel”, o dado coletado no dia D é substituído pela média dos dados coletados nos dias D-1, D-2 e D-3. Deste modo, os dados filtrados variam mais lentamente do que os dados reais, o que evidencia uma tendência média ao invés de levar em conta as variações mais bruscas causadas por erros e dados espúrios.

## 2.1.4 Objetivação

De acordo com a Heurística do Objetivo[1], foram identificados **Efeitos Adversos** e suas respectivas **Causas**.

Deve-se ressaltar que estes efeitos e suas respectivas causas foram arbitrados pelo autor e não representam a realidade da Agência Nacional de Águas. Trata-se apenas de conjecturas para permitir a resolução desta Lista de Exercícios.

Em seguida, foram propostas:

- **Tarefas** cuja execução elimina as **Causas** e
- **Propósitos** cujo atendimento elimina os efeitos adversos.

Concatenando **Tarefas** e **Propósitos** foi obtido o **Enunciado do Problema**.

Em seguida, foram analisadas algumas **Alternativas de Soluções Possíveis** (ASP), as quais foram submetidas a **Análises de Praticabilidade e Aceitabilidade** (APA).

Finalmente, foi definido o enunciado da **Alternativa de Solução Escolhida** (ASE) que é o **objetivo** procurado.

Por concisão, as “séries históricas de dados hidrológicos da Agência Nacional de Águas (ANA)” são referidas nos textos a seguir como sendo apenas “séries históricas”.

### 2.1.4.1 Definição do Problema

Considerou-se, para fins desta Lista de Exercícios, a ocorrência dos seguintes efeitos adversos:

- **EA1:** As séries históricas são armazenadas em arquivos inadequados para pesquisa.
- **EA2:** As séries históricas podem conter dados inválidos.
- **EA3:** Usuários empregam diferentes critérios para tratar dados inválidos.

#### 2.1.4.1.1 Causas (Por que está errado?)

Considerou-se, para fins desta Lista de Exercícios, que os efeitos adversos acima citados possuem respectivamente as causas:

- **C1:** Séries históricas são armazenadas em arquivos texto que devem ser lidos sequencialmente. Deste modo, qualquer pesquisa acarreta grande tempo de processamento pois é necessário ler toda a base de dados ao invés de usar técnicas mais ágeis como os arquivos de índices dos Bancos de Dados Relacionais. Como as pesquisas são muito

demoradas, os usuários evitam fazer determinadas análises ou as executam em volume menor de dados do que seria adequado, o que resulta em conclusões de baixa qualidade.

- **C2:** Dados inválidos são introduzidos por falhas no processo de coleta e transcrição dos valores medidos, tais como: falhas do equipamento de medida, falhas do operador destes equipamentos, erros de transcrição do dado enviado pelo operador para os sistemas de consolidação dos dados etc. Falta manutenção preventiva e evolutiva dos equipamentos de medida, treinamento dos operadores e uma política integrada de mitigação das falhas nos processos que ocorrem desde a medida até o armazenamento dos dados.
- **C3:** Diferentes pesquisadores ao acessar as mesmas séries históricas podem chegar a resultados distintos por terem usado critérios diferentes. Por exemplo: dados inexistentes de uma série de dados podem ser ignorados por um pesquisador e substituídos pela média dos dados vizinhos por outro pesquisador.

#### 2.1.4.1.2 Tarefa (O que, Quem, Onde e Quando se deseja realizar tal tarefa?)

A menos que haja observação explícita em contrário, as tarefas abaixo devem ser realizadas pelo aluno Antônio Magno Lima Espeschit (quem), em local de sua conveniência (onde) e em tempo hábil para permitir a entrega do **Sis ATMH** em 22/06/2009 (quando).

- **T1:** Armazenar séries históricas em Banco de Dados Relacional, provendo ferramentas que permitam ao usuário final realizar, de forma ágil e confiável, a análise de extensas séries históricas.
- **T2a:** Implementar rotinas, no sistema desenvolvido na tarefa T1, que permitam a aplicação de filtros e o cálculo de estatísticas em subconjuntos das séries históricas visando identificar dados inexistentes, possivelmente errados e *outliers*.
- **T2b:** Em tempo e local a ser definido, a ANA deve rever processos de medição, coleta, transmissão, transcrição e armazenamento de dados no intuito de mitigar a introdução de dados inválidos nas séries históricas.
- **T2c:** Em tempo e local a ser definido, a ANA deve implementar política regular de manutenção preventiva e evolutiva nos equipamentos de medida bem como de treinamento dos operadores e melhoria contínua dos processos de medição, coleta, transmissão, transcrição, depuração e armazenamento de dados.
- **T3a:** Definir critérios padronizados para tratamento de dados espúrios (faltantes ou inconsistentes) durante a realização de consultas ao Banco de Dados.
- **T3b:** Implementar no Banco de Dados os critérios definidos na tarefa T3a de modo a padronizar e agilizar seu uso.
- **T3c:** Em tempo e local a ser definido, a ANA deve validar os critérios implementados na tarefa T3b modificando-os caso necessário.
- **T4:** O software aplicativo obtido nas tarefas T1, T2a e T3b deve ser integrado aos demais módulos do sistema **Sis ATMH** de modo que o prazo de entrega deste último seja respeitado.

#### 2.1.4.1.3 Propósito (Para que se deseja realizar tal tarefa?)

Propiciar o armazenamento de séries históricas, agilizando a análise dos dados armazenados e permitindo a aplicação de filtros, a identificação de *outliers*, o cálculo de estatísticas e o tratamento padronizado de dados espúrios.

#### 2.1.4.1.4 *Enunciado do Problema*

Disponibilizar, em tempo hábil para que o sistema completo seja concluído, pelos alunos do curso CE-240 do Primeiro Período de 2009, no laboratório da FCMF e até 22/06/2009, um módulo de software integrável ao sistema Sis ATMH que possua ferramentas de análise visando propiciar o armazenamento de séries históricas, a análise ágil e confiável dos dados armazenados além de permitir a aplicação de filtros, a identificação de *outliers*, o cálculo de estatísticas e o tratamento padronizado de dados espúrios.

#### 2.1.4.2 **Definição da Solução**

##### 2.1.4.2.1 *Identificação de Alternativas de Soluções Possíveis (ASP) com os respectivos Resultados da Análise Adequabilidade, Praticabilidade e Aceitabilidade (APA)*

- **ASP1:** Adquirir no mercado uma ferramenta de software que resolva o problema proposto.
- **APA1:** Impraticável. Fator: disponibilidade. Os problemas a resolver são muito específicos, não tendo sido identificado no mercado uma ferramenta de software capaz de resolvê-lo.
  
- **ASP2:** Continuar usando o sistema atual.
- **APA2:** Inadequada. Fator: qualidade. O sistema atual não provê a agilidade e a qualidade necessárias.
  
- **ASP3:** Integrar as ferramentas de software já existentes, obtendo um sistema que resolva o problema proposto.
- **APA3:** Inaceitável. Fator: Custo não compensa o resultado. Software existente está obsoleto e os recursos necessários para integrá-los são maiores que os necessários para se criar um novo sistema.
  
- **ASP4:** Desenvolver as ferramentas de software conforme detalhado nas tarefas T1 a T4.
- **APA4:** Adequada, praticável e aceitável.

##### 2.1.4.2.2 *Enunciado da Alternativa de Solução Escolhida (ASE) que é o objetivo do projeto*

Desenvolver, em tempo hábil para que o sistema completo seja concluído, pelos alunos do curso CE-240 do Primeiro Período de 2009, no laboratório da FCMF e até 22/06/2009, um módulo de aplicativo de banco de dados integrável ao sistema Sis ATMH que possua ferramentas de análise visando propiciar o armazenamento de séries históricas, a análise ágil e confiável dos dados armazenados além de permitir a aplicação de filtros, a identificação de *outliers*, o cálculo de estatísticas e o tratamento padronizado de dados espúrios.

#### 2.1.4.2.3 Alternativas de Solução Conservadas (ASCs)

Nenhuma das demais soluções foi conservada.

#### 2.1.4.2.4 Alternativas de Solução Híbridas (ASHs)

Não foram consideradas soluções híbridas pois as soluções apresentadas são mutuamente excludentes.

### 2.1.5 Intitulação a Partir de Uma Redução de Escopo

*Módulo de Aplicação de Filtros e Cálculo de Estatísticas (AFE).*

### 2.1.6 Especificação de Requisitos

O Módulo de Aplicação de Filtros e Cálculo de Estatísticas (AFE) deve atender aos seguintes requisitos:

- Ser implementado pelo aluno Antônio Magno Lima Espescht.
- Ser integrável ao sistema Sis ATMH.
- Estar implantado no laboratório da FCMF, com qualidade e prazo adequados para que o sistema Sis ATMH esteja pronto em 22/06/2009.
- Propiciar o armazenamento de séries históricas de dados hidrológicos da Agência Nacional de Águas (ANA).
- Propiciar a aplicação de filtros tais como "média móvel" aos dados armazenados.
- Propiciar a identificação de dados inexistentes, possivelmente inválidos e *outliers*.
- Propiciar a aplicação de heurísticas previamente padronizadas para tratamento de dados inexistentes, possivelmente inválidos e *outliers*.
- Propiciar, dado um subconjunto de uma série histórica, o cálculo de quatro valores estatísticos que caracterizam estes dados: mínimo, média, máximo e desvio-padrão.

## 2.2 Heurística do Objetivo Aplicada à Proposta de Dissertação de Mestrado

Aplicou-se a **Heurística do Objetivo** [1] à Proposta de Dissertação de Mestrado “A Web Search Engine to Flight Test Data” [2].

### 2.2.1 Tematização

Mecanismo de busca aplicado a dados de ensaios em voo na Embraer.

### 2.2.2 Motivação

Conciliar a necessidade da Embraer de um mecanismo de busca especializado em dados de ensaios em voo com a oportunidade do autor defender uma tese de mestrado sobre este assunto no ITA.



### 2.2.3 Contextualização

Para vender seus produtos a Embraer precisa obter certificados de agências reguladoras como o *Federal Aviation Administration* (FAA) nos Estados Unidos e a *European Aviation Safety Agency* (EASA) na Europa.

Cada agência tem critérios específicos de certificação o que obriga a Embraer a realizar extensas campanhas de ensaios em voo. Cada campanha emprega vários protótipos os quais realizam centenas de voos no Brasil e no exterior.

Em cada voo são medidos milhares de parâmetros tais como velocidade, altitude, pressão do óleo do motor etc. ao longo de várias horas e coletando até milhares de amostras por segundo. Assim, em cada ensaio, obtém-se gigabytes de dados que devem ser rapidamente analisados pelos engenheiros para atestar o cumprimento de requisitos de certificação e para melhorar a qualidade e a segurança dos aviões fabricados pela Embraer.

O volume de dados é de tal ordem que, com as ferramentas atuais de análise, torna-se impraticável realizar a busca sistemática em toda a base à procura de uma determinada configuração de voo na qual deseja-se medir determinados parâmetros. Neste caso, a empresa opta por realizar um novo ensaio com o protótipo na configuração desejada.

Porém, o custo de cada ensaio é da ordem de dezenas de milhares de dólares por hora de voo. Portanto, se for possível evitar a realização de alguns ensaios, haverá grande economia para a empresa.

Adicionalmente, evitando-se a realização de alguns voos, a campanha de ensaios será concluída em menor tempo, permitindo que a Embraer coloque seus produtos à venda antes de seus concorrentes o que resulta em uma excelente vantagem competitiva.

### 2.2.4 Objetivação

De acordo com a Heurística do Objetivo[1], foram identificados **Efeitos Adversos** e suas respectivas **Causas**.

Deve-se ressaltar que estes efeitos, e suas respectivas causas, foram arbitrados pelo autor e não representam a realidade da Embraer. Trata-se apenas de conjecturas para permitir a resolução desta Lista de Exercícios e portanto, sua divulgação não constitui quebra de contrato quanto à confidencialidade dos procedimentos de ensaios da Embraer.

Em seguida, foram propostas:

- **Tarefas** cuja execução elimina as **Causas** e
- **Propósitos** cujo atendimento elimina os efeitos adversos.

Concatenando **Tarefas** e **Propósitos** foi obtido o **Enunciado do Problema**.

Em seguida, foram analisadas algumas **Alternativas de Soluções Possíveis** (ASP), as quais foram submetidas a **Análises de Praticabilidade e Aceitabilidade** (APA).

Finalmente, foi definido o enunciado da **Alternativa de Solução Escolhida** (ASE) que é o **objetivo** procurado.

### 2.2.4.1 Definição do Problema

Por concisão, os “dados de diversos voos das campanhas de ensaio de diversos produtos da Embraer” são referidos nos textos a seguir como sendo apenas “dados de ensaio”.

#### 2.2.4.1.1 Efeitos Adversos (*O que está errado?*)

- **EA1:** Dados de ensaio são coletados em formato binário o que agiliza a coleta e o armazenamento mas torna a consulta muito lenta e complexa.
- **EA2:** O volume de dados de ensaio é da ordem de dezenas de terabytes o que, com as ferramentas de análise atuais, inviabiliza uma pesquisa em toda a base em tempo hábil.
- **EA3:** As ferramentas de análise atuais são aplicativos desktop que só podem ser usados localmente.

#### 2.2.4.1.2 Causas (*Por que está errado?*)

- **C1:** As ferramentas de consulta são complexas e efetuam busca sequencial nos dados de ensaio. Alguns dos dados coletados têm que ser aplicados a complexas tabelas e equações, cuja definição é alterada ao longo dos ensaios, para só então resultar em valores analisáveis pela Engenharia. Logo, o processo é muito lento para ser aplicado a toda a base de dados.
- **C2:** O espaço de armazenamento é limitado. Resultados antigos são removidos para dar espaço aos voos mais recentes. Quando necessário, a restauração dos voos antigos é muito lenta. É mais rápido realizar um novo ensaio do que procurar, na base de dados de ensaios realizados, por condições iguais às que se deseja obter em um novo ensaio.
- **C3:** Dados são coletados em várias partes do mundo e as ferramentas de análise não podem ser usadas remotamente. Assim, as missões no exterior são complexas e requerem o envio de equipamentos e técnicos especializados para que estes operem as ferramentas de análise no local onde os dados de ensaio são coletados.

#### 2.2.4.1.3 Tarefa (*O que, Quem, Onde e Quando se deseja realizar tal tarefa?*)

As tarefas abaixo devem ser realizadas pelo aluno Antônio Magno Lima Espeschit (quem), em local de sua conveniência (onde) e em tempo hábil de modo permitir a entrega do sistema em condições de uso por usuário final da Embraer até dezembro de 2010 (quando).

- **T1:** Criar uma ferramenta de software capaz de ler dados de ensaio em formato binário, aplicar equações de obtenção de parâmetros derivados e armazenar os resultados em formato mais adequado à consulta.
- **T2a:** Criar uma ferramenta de software capaz de comprimir os dados obtidos na tarefa T1, mesmo que haja perdas de precisão neste processo, reduzindo de tal forma o volume de dados que se torne desnecessário remover os dados de ensaios antigos para dar espaço aos dados de ensaios mais recentes.
- **T2b:** Criar um aplicativo de software de banco de dados capaz de armazenar, em um Banco de Dados Relacional, os dados comprimidos obtidos na tarefa T2a e de permitir acesso a estes dados via SQL (*Structured Query Language*).
- **T3:** Criar uma interface web que, usando uma linguagem específica para dados de ensaios em voo, permita realizar consultas no aplicativo criado na tarefa T2b, viabilizando

pesquisas, locais e via Internet, em toda a base de dados de ensaio, retornando os resultados mais relevantes em apenas alguns segundos.

- **T4:** Implantar na Embraer os sistemas desenvolvidos nas tarefas acima.

Espera-se que os erros de precisão da tarefa T2a não prejudiquem a qualidade dos dados consultados. Porém, após ter usado o engenho de busca para identificar os trechos dos ensaios já realizados, pode-se usar o sistema atual para rapidamente obter os valores com máxima precisão.

#### *2.2.4.1.4 Propósito (Para que se deseja realizar tal tarefa?)*

Permitir a consulta, local ou via Internet, de toda a base de ensaios em apenas alguns segundos de modo a evitar a realização de um novo ensaio quando for possível extrair os mesmos resultados a partir dos dados de um ensaio já realizado, reduzindo custos e agilizando a campanha de ensaios.

#### *2.2.4.1.5 Enunciado do Problema*

Armazenar, em um Banco de Dados Relacional, uma versão comprimida e menos precisa dos dados de ensaio, permitindo consultar toda a campanha de ensaios em poucos minutos, numa interface web, localmente ou via Internet, visando evitar a realização de um novo ensaio quando for possível extrair os mesmos resultados a partir dos dados de um ensaio já realizado, reduzindo custos e agilizando a campanha de ensaios.

### **2.2.4.2 Definição da Solução**

#### *2.2.4.2.1 Identificação de Alternativas de Soluções Possíveis (ASP) com os respectivos Resultados da Análise Adequabilidade, Praticabilidade e Aceitabilidade (APA)*

- **ASP1:** Integrar ferramentas de software já existentes, obtendo um sistema que atenda aos requisitos do problema.
- **APA1:** Impraticável. Fator: disponibilidade. Sistemas atuais realizam busca sequencial que é muito lenta para o problema proposto.
- **ASP2:** Desenvolver os aplicativos conforme detalhado nas tarefas T1 a T4, realizando compressão dos dados do seguinte modo: dividir o ensaio em intervalos de tempo regulares, armazenando valores de média e desvio padrão para cada intervalo.
- **APA2:** Adequada, praticável e aceitável.
- **ASP3:** Idêntica a ASP2 porém realizando compressão dos dados definindo Segmentos Homogêneos [2] para os quais se calculam valores de média e desvio padrão.
- **APA3:** Adequada, praticável e aceitável. Resulta em mais tempo de processamento que a ASP2 porém menor espaço de armazenamento e maior agilidade de consulta.

- **ASP4:** Idêntica a ASP3 porém realizando a definição de Segmentos Homogêneos usando algoritmo Bayesiano [3] que é sensível a variações de frequência e amplitude, ao invés de apenas amplitude como ocorre na ASP3.
- **APA4:** Adequada, praticável e aceitável. Resulta em mais tempo de processamento que a ASP3 porém menor espaço de armazenamento e maior agilidade de consulta.

Após análises, a Alternativa de Solução Escolhida (ASE) foi a ASP4.

#### 2.2.4.2.2 *Enunciado da Alternativa de Solução Escolhida (ASE) que é o objetivo do projeto*

Desenvolver, até 20/12/2010, uma ferramenta de software aplicativo capaz de ler e processar dados de ensaio, comprimindo-os com *Segmentos Homogêneos Bayesianos* [3] em um Banco de Dados Relacional consultável por uma interface web usando uma linguagem específica para consulta a dados de ensaio, visando propiciar a consulta, local ou via Internet, de toda a base de ensaios em no máximo alguns minutos de modo a evitar a realização de um novo ensaio quando for possível extrair os mesmos resultados a partir dos dados de um ensaio já realizado, reduzindo custos e agilizando a campanha de ensaios da Embraer.

#### 2.2.4.2.3 *Alternativas de Solução Conservadas (ASCs)*

A alternativa ASP1 foi descartada e as demais conservadas.

#### 2.2.4.2.4 *Alternativas de Solução Híbridas (ASHs)*

Considerou-se que a ASP4 é uma solução híbrida pois as alternativas ASP2 e ASP3 constituem passos intermediários em direção à alternativa ASP4.

### 2.2.5 **Intitulação a Partir de Uma Redução de Escopo**

**A Web Search Engine to Flight Test Data Using Homogeneous Segments.**

### 2.2.6 **Especificação de Requisitos**

O Mecanismo de Busca com Interface Web para Dados de Ensaios em Voo Usando Segmentos Homogêneos deve propiciar:

- Leitura de dados de ensaio em formato binário obtendo valores de parâmetros primários tais como velocidade, pressão, temperatura etc. e sua variação com o tempo.
- Cálculo de parâmetros derivados tais como “Número de Mach” [7] de acordo com tabelas e equações fornecidas pela Engenharia de Desenvolvimento da Embraer.
- Armazenamento de parâmetros primários e derivados em arquivos texto.
- Compressão dos dados dos arquivos texto, usando *Segmentos Homogêneos Bayesianos* [3], com armazenamento dos dados comprimidos em um Banco de Dados Relacional.
- Consulta, local ou via Internet, via interface web, dos dados armazenados no Banco de Dados Relacional, usando linguagem específica para dados de ensaios em voo.
- Obtenção de resultados em menos de 1 minuto, mesmo que o escopo inclua até 10 parâmetros em todos os voos em todos os protótipos de um determinado produto.

- A partir dos resultados encontrados no sistema, efetuar a criação automática de arquivo de consulta nos sistemas atuais, visando obter dados com maior precisão que a disponível nos dados comprimidos.
- Redução de custos pela utilização dos dados de ensaios já realizados ao invés da realização de um novo ensaio.
- Aumento da vantagem competitiva da Embraer face seus concorrentes pelo encurtamento da campanha de ensaios.

## 2.3 Pesquisas na Internet

Realizou-se pesquisas na internet usando os mecanismos de busca Google [4] e CiteSeerX[6].

### 2.3.1 Pesquisa sobre Filtros e Cálculos Estatísticos para Séries Hidrológicas (Temática de Protótipo de Projeto de Aplicativo de Banco de Dados)

Principais artigos encontrados:

- AGÊNCIA NACIONAL DE ÁGUAS (ANA). *Caderno de Recursos Hídricos*. Disponível em: <http://www.ana.gov.br/sprtew/recursosohidricos.asp>. Acessado em: 23 de março de 2009.
- CARVALHO, LEILA MARIA VÉSPOLI DE. *Filtragem de Séries Temporais*. Disponível em: [http://www.icesb.ucsb.edu/gem/material\\_filtros/filtragem.doc](http://www.icesb.ucsb.edu/gem/material_filtros/filtragem.doc). Acessado em: 23 de março de 2009.
- NASCIMENTO, NILO DE OLIVEIRA. *Avaliação de Tendências em Séries Temporais*. Disponível em: <http://www.etg.ufmg.br/tim1/tendencia-temp.doc>. Acessado em: 23 de março de 2009.
- GONZÁLEZ, MARCO AURÉLIO STUMPF. *Outliers*. Disponível em: <http://www.exatec.unisinos.br/~gonzalez/valor/inferenc/pressup/outliers.html>. Acessado em: 23 de março de 2009.
- COLLISCHONN, WALTER **Simulação Hidrológica de Grandes Bacias**. Tese de Doutorado em Engenharia de Recursos Hídricos e Saneamento Ambiental da Universidade Federal do Rio Grande do Sul (UFRS) 2001, Disponível em: [galileu.iph.ufrgs.br/collischonn/climarh/download/tese%20Walter%20Collischonn.pdf](http://galileu.iph.ufrgs.br/collischonn/climarh/download/tese%20Walter%20Collischonn.pdf). Acessado em: 23 de março de 2009.
- TUCCI, CARLOS. *Métodos de Análises de Alterações de Séries Hidrológicas*. Disponível em: <http://blog.rhama.net/2008/09/14/alteracao-das-series-hidrologicas/>. Acessado em: 23 de março de 2009.

### 2.3.2 Pesquisa sobre Mecanismos de Busca para Parâmetros de Ensaio (Temática de Mestrado)

Principais artigos encontrados:

- CUBBINS, THOMAS D., HOPE, THOMAS R., MICHELS, JOHN M. Data compression method for telemetry of vibration data. Patente 4908785. Disponível em: <http://www.freepatentsonline.com/4908785.html>. Acessado em: 23 de março de 2009.

- FRIDTJOF, THOMAS. **Automated Road Segmentation using a Bayesian Algorithm** *Journal of Transportation Engineering*, Vol. 131, No. 8, August 2005, pp. 591-598. Disponível em: <http://cedb.asce.org/cgi/WWWdisplay.cgi?0524137>. Acessado em: 23 de março de 2009.
- NI, KAI; PAISLEY, JOHN; CARIN, LAWRENCE; DUNSON, DAVID **Multi-Task Learning for Analyzing and Sorting Large Databases of Sequential Data** (2008). Disponível em: [http://www.ece.duke.edu/~lcarin/MTL\\_Sequential\\_Data\\_Revised.pdf](http://www.ece.duke.edu/~lcarin/MTL_Sequential_Data_Revised.pdf). Acessado em: 24 de março de 2009.
- SARNOWSKI, WOJCIECH; SZAJOWSKI, KRZYSZTOF. **Optimal detection of homogeneous segment of observations in stochastic sequence.** Disponível em: <http://adsabs.harvard.edu/abs/2008arXiv0812.3632S>. Acessado em: 23 de março de 2009.

### 2.3.3 Pesquisa sobre Genealogia (Hobby)

Principais artigos encontrados:

- ALMEIDA, VIRGILIO PEREIRA DE. *Nossa Gente Genealogia*. Disponível em: <http://www.nggenealogia.com.br>. Acessado em: 28 de março de 2009.
- ARAÚJO, MAGNO JOSÉ DE SÁ. *Genealogia Pernambucana*. Disponível em: <http://www.araujo.eti.br/familia.asp?numPessoa=33628>. Acessado em: 28 de março de 2009.
- VILLAS BÔAS, LUÍS ANTÔNIO. *GeneaMinas – Genealogia e História de Minas Gerais*. Disponível em: <http://www.geneaminas.com.br>. Acessado em: 28 de março de 2009.
- ESPESCHIT, ANTÔNIO MAGNO LIMA. *Genealogia da Família Espeschit*. Disponível em: <http://www.espeschit.com.br>. Acessado em: 23 de março de 2009.
- THE CHURCH OF JESUS CHRIST OF LATTER-DAY SAINTS. *Family Search*. Disponível em: <http://www.familysearch.org/eng/default.asp>. Acessado em: 23 de março de 2009.

## 3 Conclusão e Comentários

O uso da Heurística do Objetivo [1] nos levou a pensar, de forma metódica e ordenada, a respeito dos problemas que pretendemos resolver e das possíveis soluções.

Os objetivos que definimos com esta técnica se mostraram mais sucintos e bem formados que aqueles obtidos de forma “intuitiva”, sem empregar a heurística.

O poder da heurística ficou bastante evidente quando a apliquei em minha proposta de tese de mestrado. A técnica, aplicada em apenas algumas horas, elucidou aspectos que eu não havia ainda abordado, apesar de ter empregado meses na criação da proposta.

É interessante observar que a heurística não pode ser aplicada de forma simples e direta pois é comum que o detalhamento em uma determinada etapa nos leve a rever hipóteses ou diagnósticos feitos em etapas anteriores. Porém, o processo segue inexoravelmente em direção à definição de um objetivo que é ao mesmo tempo econômico na quantidade de palavras e completo na abrangência dos detalhes que realmente importam na solução do problema.

## 4 Referências

- [1] CUNHA, ADILSON MARQUES DA. Notas de Aula de “CE-240 Projeto de Sistemas de Bancos de Dados” no Primeiro Período de 2009. ITA – Instituto Tecnológico de Aeronáutica. Disponível em: <http://www.comp.ita.br/~cunha>. Acessado em: 23 de março de 2009.
- [2] ESPESCHIT, ANTÔNIO MAGNO LIMA. **A Web Search Engine to Flight Test Data**. M. Sc. Proposal 2009. ITA – Instituto Tecnológico de Aeronáutica. Disponível em: [http://ita90.com.br/~toni/AESPESCHIT\\_mestrado.pdf](http://ita90.com.br/~toni/AESPESCHIT_mestrado.pdf). Acessado em: 23 de março de 2009.
- [3] FRIDTJOF, THOMAS. **Generating homogeneous road sections based on surface measurements: available methods**. 2nd European Pavement and Asset Management Conference, 21-23 March 2004.
- [4] GOOGLE INC. Engenho de Busca na Internet. 2009. Disponível em: <http://www.google.com>. Acessado em: 23 de março de 2009.
- [5] HUFF, DARREL. **How To Lie with Statistics**. W. W. Norton & Company. New York. 1993. ISBN 0-393-31072-8.
- [6] THE COLLEGE OF INFORMATION SCIENCES AND TECHNOLOGY AT PENN STATE. CiteSeerX Disponível em: <http://citeseerx.ist.psu.edu>. 2009. Acessado em: 23 de março de 2009.
- [7] WIKIMEDIA FOUNDATION, INC. Wikipedia (*Número de Mach*). Disponível em: [http://pt.wikipedia.org/wiki/Número\\_de\\_Mach](http://pt.wikipedia.org/wiki/Número_de_Mach). 2009. Visitado em: 28 de março de 2009.
- [8] WIKIMEDIA FOUNDATION, INC. Wikipedia (*Outlier*). Disponível em: <http://en.wikipedia.org/wiki/Outlier>. 2009. Visitado em: 28 de março de 2009.