

## **DESENVOLVIMENTO DE INFRAESTRUTURA E FERRAMENTAS PARA APLICAÇÕES DE CITIZEN SCIENCE**

### **RELATÓRIO FINAL DE PROJETO DE INICIAÇÃO CIENTÍFICA (PIBIC/CNPq/INPE)**

Danilo Clemente Briante (UNIFESP, Bolsista, PIBIC/CNPq)  
E-mail: [danilo.clemente@unifesp.br](mailto:danilo.clemente@unifesp.br)

Nandamudi Lankalapalli Vijaykumar (LAC/CTE/INPE, Orientador)  
E-mail: [vijay@lac.inpe.br](mailto:vijay@lac.inpe.br)

### **COLABORADORES**

Dr. Luciano Godoy Fagundes (Colaborador/INPE)

Julho de 2011

## **SUMÁRIO:**

### **CAPÍTULO 1 – RESUMO**

### **CAPÍTULO 2 – INTRODUÇÃO**

### **CAPÍTULO 3 – PROJETO**

3.1 Ferramentas

3.2 JAVA

3.3 NetBeans

3.4 Iphone

3.5 TomCat

3.6 HSQLDB

### **CAPÍTULO 4 – IMPLEMENTAÇÃO**

4.1 Comunicação

4.1.2 FirstLoad

4.1.3 Update

4.1.4 *Extensible Markup Language*

4.2 Busca de Notícias

4.2.1 WebSpiders

4.3 Banco de Dados

4.3.1 Inserção de noticias

4.3.2 Remoção de noticias

### **CAPÍTULO 5 – INTEGRAÇÃO COM REDES SOCIAIS**

### **CAPÍTULO 6 – APLICAÇÃO DE CITIZEN SCIENCE**

### **CAPÍTULO 7 – TRABALHOS FUTUROS**

### **CAPÍTULO 8 – CONCLUSÃO**

### **BIBLIOGRAFIA**

## CAPITULO 1 – RESUMO

Este trabalho, iniciado em agosto de 2010 tem como objetivo elaborar uma revista digital para dispositivos móveis como prova de conceito de técnicas de *citizen science* e *outreach*. A revista serviria para divulgar notícias, trabalhos e atividades realizadas no INPE entre a comunidade, despertando assim o maior interesse pela instituição. A princípio foram realizados treinamentos para que os alunos pudessem receber uma introdução a linguagem Objective-C, linguagem esta utilizada para desenvolver aplicativos para iPhone. Para demonstrar a viabilidade de usar técnicas de *citizen science* foi definido um protótipo de jogo iterativo chamado “pinte a floresta”, onde voluntários poderiam fazer tarefas simples de interpretação de imagens digitais, possibilitando assim à população participar de projetos científicos sem a necessidade de um treinamento específico. Outra técnica aplicada foi o conceito de *outreach*, o qual baseia-se em divulgar informações para tal, foi criado um servidor TomCat<sup>1</sup>, onde aplicamos o conceito de Web Crawler<sup>2</sup> varrendo o site do INPE<sup>3</sup> em busca de notícias cadastradas recentemente, essas são inseridas no banco de dados (o banco de dados utilizado foi o HSQLDB<sup>4</sup>). O envio de notícias para a aplicação é dado através da troca de informações do servidor com a aplicação, quando realizada essa chamada o servidor carrega todas as informações mais recentes no banco de dados e realiza um parser codificando através do conceito de XML<sup>5</sup> as notícias e enviando-as para a aplicação via Internet, onde lá elas serão organizadas e exibidas ao usuário.

---

<sup>1</sup> <http://tomcat.apache.org/>

<sup>2</sup> [http://en.wikipedia.org/wiki/Web\\_crawler](http://en.wikipedia.org/wiki/Web_crawler)

<sup>3</sup> <http://www.inpe.br/>

<sup>7</sup> <http://hsqldb.org/>

<sup>5</sup> <http://pt.wikipedia.org/wiki/XML>

## CAPÍTULO 2 – INTRODUÇÃO

Este projeto tem como objetivo aplicar os conceitos de *citizen science* e *outreach* de maneira a fornecer uma infra-estrutura para aplicações móveis, para que seja realizada uma ampla divulgação das notícias e atividades realizadas dentro do Instituto Nacional de Pesquisas Espaciais (INPE) possibilitando um maior envolvimento entre o instituto e a comunidade.

O conceito de *Outreach*, envolve atividades de divulgação, nas quais são apresentados trabalhos, idéias e atividades de forma acessível e educativa, a fim de disseminar as informações possibilitando assim, aumentar o interesse da comunidade pela instituição.

Por outro lado, o conceito de *Citizen Science* é baseado em atividades onde a população participa em projetos científicos para resolver um determinado problema sem a necessidade de treinamento científico.

Dessa forma, o projeto baseia-se em criar uma revista digital para dispositivos móveis como prova do conceito das duas técnicas.

A revista digital abordaria o conceito de *outreach* conforme explicado anteriormente, fornecendo ao usuário notícias e informações do INPE, além de permitir a utilização do conceito de *citizen science* possibilitando o usuário interagir com o software de maneira a produzir uma informação desejada, no caso foi implementado jogos interativos como, por exemplo, um pinte a floresta, jogo no qual o usuário detecta a presença de uma floresta colorindo a região.

Para tal, torna-se necessário à elaboração de uma infra-estrutura que possibilite o armazenamento e envio de notícias, tal como uma estrutura que possibilite a busca automatizada das mesmas em um ambiente fornecedor podendo armazená-las dentro de um banco de dados, tornando assim mais fácil e rápido o fornecimento das notícias.

## CAPÍTULO 3 – PROJETO

### 3.1 Ferramentas:

Para a implementação do projeto, foram definidas linguagens de programação na qual, por parte do servidor foi selecionada a linguagem JAVA<sup>6</sup>, onde foi escolhida a plataforma NetBeans<sup>7</sup> como auxílio na implementação. Por parte do cliente, ou seja, a aplicação móvel, foi escolhido o aparelho iPhone<sup>8</sup>, com o sistema IOs (sistema operacional do iPhone), utilizando a devida linguagem de programação necessária para se criar aplicativos para o mesmo, Objective-C<sup>9</sup>.

Outra ferramenta utilizada por parte do servidor foi o Tomcat<sup>10</sup>, aplicativo que atua como um servidor permitindo assim a comunicação entre o servidor e a aplicação. Para o armazenamento dos dados coletados, tornou-se necessário à utilização de um servidor de banco de dados, o HSQLDB<sup>11</sup>, o qual foi escolhido devido a grande compatibilidade com a linguagem Java e também com a plataforma de trabalho (NetBeans).

Além dessas ferramentas foi necessário à alocação de um espaço dentro de um servidor situado no INPE onde as informações seriam armazenadas e distribuídas para outros usuários, necessitando permanecer ligado 24 horas por dia, para que não houvesse problemas de comunicação do aplicativo.

### 3.2 JAVA:

A linguagem JAVA é uma linguagem de alto nível, orientada a objetos desenvolvida na década de 90 por uma equipe de programadores na empresa Sun Microsystems<sup>12</sup>, que

---

<sup>6</sup> [http://www.java.com/pt\\_BR/](http://www.java.com/pt_BR/)

<sup>7</sup> <http://netbeans.org/>

<sup>8</sup> <http://www.apple.com/iphone/>

<sup>9</sup> <http://pt.wikipedia.org/wiki/Objective-C>

<sup>10</sup> <http://tomcat.apache.org/>

<sup>11</sup> <http://hsqldb.org/>

<sup>12</sup> [http://pt.wikipedia.org/wiki/Sun\\_Microsystems](http://pt.wikipedia.org/wiki/Sun_Microsystems)

diferente das outras linguagens o código é compilado para ByteCode<sup>13</sup> onde é executado por uma máquina virtual, possibilitando assim ser uma linguagem extremamente portátil. Uma das maiores características da linguagem é a sua portabilidade e sua facilidade de programar por ser uma linguagem de alto nível.

### **3.3 NetBeans:**

NetBeans é um Ambiente de Desenvolvimento Integrado (IDE) gratuito e de código aberto que foi desenvolvido inicialmente em 1996 por estudantes da Universidade de Charles<sup>14</sup>. Entre suas maiores vantagens o suporte a diversas linguagens torna-se um item de referência.

### **3.4 Iphone:**

O Iphone é um smartphone desenvolvido pela Apple apresentado em 2007, dentro das vantagens é interessante destacar que os aparelhos Iphone possuem características presentes em todos os modelos tornando possível a utilização de um aplicativo em todos os modelos diferentemente de outros aparelhos como por exemplo, os aparelhos com sistema Android<sup>15</sup>, sistema desenvolvido pela Google<sup>16</sup> que pode rodar em aparelhos sem que seja necessário possuir todos os recursos por exemplo, um celular pode possuir o sistema Android e não possuir câmera ou recursos de sistema de posicionamento global (GPS), enquanto outro aparelho pode possuir esses recursos, podendo fazer com que aplicativos possam ser inutilizados pelo aparelho por falta de recurso.

---

<sup>13</sup> <http://pt.wikipedia.org/wiki/Bytecode>

<sup>14</sup> <http://www.cuni.cz/UKENG-1.html>

<sup>15</sup> <http://www.android.com/>

<sup>16</sup> <http://www.google.com/>

### **3.5 TomCat:**

TomCat é software livre de um servidor WebJava desenvolvido pela Apache Software Foundation, dentro do projeto Apache Jakarta<sup>17</sup>.

### **3.6 HSQLDB:**

O HSQLDB ou Hyperthreaded Structured Query Language Database, é um sistema de gerenciamento de banco de dados de código livre desenvolvido em 1998 e escrito totalmente em linguagem Java, possuindo um bom desempenho.

## CAPÍTULO 4 - IMPLEMENTAÇÃO

### 4.1 Comunicação:

A configuração correta do TomCat permite a comunicação do servidor com a aplicação via Internet. Tal comunicação é realizada através da abertura de uma porta no próprio servidor.

Em Java a comunicação é dada por meio de servlets<sup>18</sup>. No contexto do projeto a comunicação é dada através de requisições da revista, onde cada requisição envia para o servidor o UID<sup>19</sup> do aparelho e uma função que se deseja executar. As funções foram separadas em *FirstLoad* e *update*.

#### 4.1.2 FirstLoad:

Nessa função o servidor armazena no banco de dados o UID do usuário permitindo assim um cadastro de usuários no aplicativo, logo em seguida o servidor retorna como resposta as notícias atualizadas que estão no banco de dados.

#### 4.1.3 Update:

Nessa função o servidor seleciona no banco de dados as notícias mais recentes e as envia para o smartphone.

#### 4.1.4 Extensible Markup Language:

Mais conhecido como XML<sup>20</sup> essa linguagem é muito utilizada em comunicação via internet por se tratar de uma forma simples e legível de se separar conteúdo através de *tags*. Essa linguagem foi adotada para realizar a transferência das notícias do servidor para a aplicação.

---

<sup>18</sup> Classe na linguagem Java que processa requisições e respostas.

<sup>19</sup> Código que permite a diferenciação de aparelho, muito semelhante ao código MAC do computador.

<sup>20</sup> <http://pt.wikipedia.org/wiki/XML>

## **4.2 Busca de Notícias:**

### **4.2.1 WebSpider:**

O sistema de busca é baseado em agentes chamados de WebSpiders.

WebSpiders são agentes automatizados que percorrem sites na internet de um modo específico, com um objetivo específico, tais agentes também são utilizados em grandes sites de busca como o Google.

No caso do servidor, os webspiders varrem o site do Instituto em busca de notícias, cadastrando-as no banco de dados.

Inicialmente o servidor cria uma *thread*<sup>21</sup> que atuará como *Spider* no site do INPE realizando um parser no endereço de notícias gravando em uma tabela os endereços de notícias ainda não visitadas. Em seguida são criadas threads que visitarão esses endereços em busca da notícia, armazenando as informações (o armazenamento será definido na seção 3.3) no banco de dados. Após o cadastro a thread da notícia é eliminada Este ciclo se repete com um tempo pequeno de espera enquanto o servidor estiver ligado, podendo assim, manter o banco de dados sempre atualizado.

### **4.3 Banco de Dados:**

O banco de dados gerencia as informações das notícias, além de informações dos usuários. Por parte do usuário o banco de dados armazena somente o UID, já definido anteriormente. Por parte das notícias o banco de dados armazena informações como:

- Identificação da notícia
- Título;
- Descrição;
- Foto;
- Texto;

---

<sup>21</sup> [http://pt.wikipedia.org/wiki/Thread\\_\(ci%C3%AAncia\\_da\\_computa%C3%A7%C3%A3o\)](http://pt.wikipedia.org/wiki/Thread_(ci%C3%AAncia_da_computa%C3%A7%C3%A3o))

- Autor;
- Tags;
- Views;
- Data da notícia.

Dentro desses dados, é importante definir alguns não triviais.

Identificação da notícia: Número gerado automaticamente no momento de cadastro permitindo ao Banco de Dados identificar a notícia.

Foto: armazena uma foto da notícia caso a mesma possua uma imagem.

Autor: armazena o autor da noticia caso haja, caso contrario armazena como sendo INPE o autor.

Tags: Palavras chaves que definem o texto.

Views: armazena o número de visualizações da noticia.

#### **4.3.1 Inserção de notícias:**

Após realizar o parser das informações as notícias são enviadas para o banco de dados, onde ele verifica se já existe a notícia encontrada, caso ele já possua a notícia a função retorna um valor definido para thread onde lá ela será finalizada, caso contrario é atribuído uma identificação para a notícia e a mesma é armazenada no banco de dados.

#### **4.3.2 Remoção de notícias:**

A cada tempo pré-determinado é realizada uma varredura pelo banco de dados onde as notícias antigas são removidas possibilitando o banco armazenar outras noticias.

## CAPÍTULO 5 INTEGRAÇÃO COM REDES SOCIAIS

Atualmente a maior parte dos aplicativos possuem integração com redes sociais, portanto tornou-se interessante a integração do aplicativo com algumas redes sociais, como por exemplo o aplicativo realiza uma integração com o Twitter<sup>22</sup>, onde o usuário pode postar notícias sobre a região onde está realizando um serviço de *citizen science* pode informar através de sua conta um buraco na rua sendo que a posição é adquirida através da própria aplicação através do Sistema de Posicionamento Global (GPS).

---

<sup>22</sup> <http://twitter.com>

## CAPÍTULO 6 – APLICAÇÃO DE CITIZEN SCIENCE

A aplicação de técnicas de *citizen science* é dada através de sub-aplicativos dentro da revista digital, onde é oferecidas ao usuário uma interação com a aplicação, um dos aplicativos disponível é o pinte a floresta. Como o próprio nome já diz o usuário pinta regiões de forma interativa possibilitando o envio das informações para o servidor onde podem ser analisadas conforme o interesse do administrador da aplicação.

Outra forma de *citizen science* é, conforme dito na seção 4, a possibilidade de divulgar informações através da rede social Twitter.

## CAPÍTULO 7 – TRABALHOS FUTUROS

Alguns trabalhos que podem surgir a partir desse projeto são descritos a baixo:

- Integração da aplicação com outras redes sociais;
- Mineração dos dados dentro do servidor para possíveis análise;
- Aumento do numero de aplicativos de *citizen science* dentro da revista;
- Coleta de outros tipos de informação do usuário como fotos;
- Ampliação do numero de sites em que são buscadas as notícias.

## CAPÍTULO 8 – CONCLUSÃO

Como prova de conceito, tornou-se possível a elaboração da revista digital em um sistema móvel, onde são aplicadas técnicas de *outreach* divulgando assim as atividades e notícias do INPE e por meio de técnicas de *citizen science* por parte da aplicação são coletados dados de maneira a futuramente serem refinados através de uma mineração de dados de maneira que possa utilizar os dados para obter um resultado esperado.

## **BIBLIOGRAFIA**

[1]. Site do INPE: [www.inpe.br](http://www.inpe.br)

[2]. Tutorial de instalação do Apache TomCat:

<http://www.mhavila.com.br/topicos/java/tomcat.html>

[3]. Tutorial de instalação HSQLDB no NetBeans:

<http://www.guj.com.br/java/81539-tutorial-de-hsqldb-e-java-para-netbeans-e-eclipse>