

# **Crawler e Ferramenta de Visualização de Dados**

## **WebCrawler Kaizoku**

**Caio M. Viera Silva, Denis L. Oliveira, Gustavo B. Oliveira**

Centro Universitário SENAC – Unidade Santo Amaro  
CEP 04750-000 - São Paulo – SP – Brasil

caiomvs@hotmail.com, denis\_blackshadow@hotmail.com,  
gubaroli@hotmail.com

**Abstract.** *The goal of the project is to develop a web crawler, a program made in python in order to get links from one web page, presenting visually to the user your data previously classified into themes.*

*Through a url provided by the user, the crawler will catalog and classify information of the site in question, the result will be visually presented to the user on a page created by the crawler using html5 and css3.*

*keys words: Web Crawler, Integrate project, Data Visual*

**Resumo.** *O projeto se trata de desenvolver um web crawler, um programa feito em python com a finalidade de buscar os links de uma página web, apresentando visualmente para o usuário os seus dados previamente classificados em temas.*

*Através de uma url fornecida pelo usuário, o crawler vai catalogar e classificar as informações do site em questão, o resultado será apresentado visualmente ao usuário em uma página criada pelo crawler utilizando html5 e css3.*

*Palavras chaves: Web Crawler, Projeto Intregador, Visualização de Dados*

## **Introdução**

Este projeto tem por meio desenvolver e integrar uma aplicação de um web crawler gerado a partir da linguagem de programação Python, e uma ferramenta de visualização de dados aplicada a um browser de internet, produzida com a utilização de HTML 5, CSS3 e Java Script.

Onde essa integração resultará numa pagina onde o “usuário” fornece um endereço web para a pagina e o crawler percorrera e irar fornecer os links e arquivos na pagina organizados em temas como por exemplos Esportes, Variedades, Informática, etc.

## WebCrawler Kaizoku

O projeto se concentra em três partes principais que são, o Web Crawler, a Ferramenta de Visualização de Dados e a Integração dos dois.

### 1.Web Crawler

Atraves da linguagem de programação Python na sua versão 3.X, desenvolvemos um web crawler utilizando as bibliotecas *xi,xi*. A função básica do crawler e basicamente percorrer o código fonte do endereço de web fornecido e identificar todos os links dele.

Logo depois dessa operação ele vai organizar a partir de temas já incluídos e classificar os links obtidos no processo de “*spidering*”. Com os dados classificados o crawler vai passar essa informação através de um servidor programado usando a biblioteca Python CherryPy as informações para a interface web que vai organizar visualmente os dados.

### 2.Ferramenta de Visualização de Dados

Desenvolvemos uma aplicação usando uma pagina desenvolvida com HTML 5, CSS3 e Java Script. Cujo pagina fornece o caminho web para o crawler, ( figura 1) e recebe a lista de links classificada e organiza num modelo de dados, “*a decidir*”(figura 2).

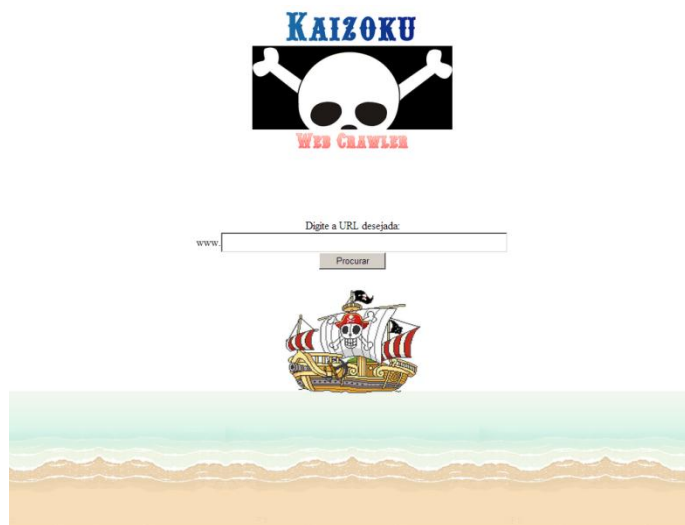


Figura 1. Tela de entrada

### **3. Interação Crawler e Web**

Através da biblioteca de Python CherryPy se tornou possível integrar o código em Python do web crawler e da página HTML. Em suma o Cherrypy funciona como um framework HTTP, que faz tudo o que é necessário para permitir que o código Python seja executado quando uma URL é solicitada pelo usuário.

Ele irá enviar o caminho da página solicitada ao código Python que irá fazer o processo de “*spidering*” e Catalogação, e fornecer ao código da página web os dados necessários para gerar a parte visual dos dados

## **Conclusão**

Web crawler Kaizoku foi um projeto que visou a união entre linguagens de programação distintas, neste caso, dependentes entre si para a execução do mesmo. Desta forma, este grupo desenvolveu um programa em Python com a finalidade identificar links e imagens contidos em um endereço web fornecido pelo usuário, cujo o resultado obtido foi mostrado ao usuário através de uma interface web.

O grupo optou escolher o tema e o nome kaizoku (que significa pirata em japonês), pois crawler é uma ferramenta que navega em uma página web e retorna os links dessa página ao usuário, o que é semelhante à história dos piratas que navegavam pelos mares e saqueavam todos os tesouros contidos neles.

## Referencias Bibliograficas

W3SCHOOLS. **JavaScript Tutorial**. Disponível em: <  
<http://www.w3schools.com/js/default.asp>>. Acesso em: 05/setembro/2012.

SCRIPTBRASIL. **Códigos Javascript**. Disponível em: <  
<http://www.scriptbrasil.com.br/codigos/javascript/>>. Acesso em: 05/setembro/2012.

ANSWERS. **How to build a simple crawler**. Disponível em:<  
<http://answers.oreilly.com/topic/1088-how-to-build-a-simple-web-crawler/>>. Acesso em:  
10/agosto/2012.

SLIDESHARE. **Desenvolvendo uma aplicação web usando o Cherrypy**. Disponível em: <  
<http://www.slideshare.net/asergionogueira/desenvolvendo-uma-aplicao-web-usando-o-cherry>  
>. Acesso em: 10/outubro/2012.

GitHub URL:

<https://github.com/denisbs/vongola>