

## # PROJETO WEB BOT

1. **Título:** DANZO - Mapeador de Criminalidade ao Redor da FATEC.

2. **Objetivo:** Raspagem de dados para demonstração da situação atual e tendência dos índices de criminalidade num certo raio da Faculdade de Tecnologia de São Paulo – Prof. Jessen Vidal.

### 3. Introdução.

Um robô da Internet ou *Internet Bot*, ou simplesmente *Web bot*, consiste num programa ou conjunto de instruções em código que verifica a teia mundial (W.W.W.) de forma metódica ou automatizada.

Este processo tem no seu cerne uma função básica chamada de *Web crawling* ou *spidering*.<sup>1</sup>

A literatura atual tende a identificar o *Web crawler* e o *Web bot* como sinônimos. Para fins deste projeto, o *web bot* [doravante denominado *WB*] consiste num robô da Internet que vasculha por dados determinados nas instruções do código de programação e geral certos resultados que tenham valor para o cliente (*product owner*).<sup>2</sup>

O WB podem ser utilizados para copiar dados de páginas da Internet que se queira vasculhar, verificando links e validar códigos HTML. Além disso, podem ser utilizados para obter dados especificados previamente, tal qual endereços de e-mail (geralmente para que se crie uma lista de SPAM), ou dados estatísticos em geral.<sup>3</sup>

Com uma variedade virtualmente infindável de usos, os WB podem ser utilizados tanto para ferramentas de suporte à automação de tarefas com objetivos benignos como para prejudicar pessoas, variando desde serviços de atendimento ao consumidor (chat bots), passando por monitoramento de informações esparsas sem controle ou repositório oficial até mesmo ataques cibernéticos por civis e militares (DDOS - *distributed denial of service attack*, ou ataques distribuídos de negação de

---

<sup>1</sup> Sience Daily. Web crawler. Em WWW <[https://www.sciencedaily.com/terms/web\\_crawler.htm](https://www.sciencedaily.com/terms/web_crawler.htm)>, acesso em 05.set.19.

<sup>2</sup> Sience Daily. Web crawler. Em WWW <[https://www.sciencedaily.com/terms/web\\_crawler.htm](https://www.sciencedaily.com/terms/web_crawler.htm)>, acesso em 05.set.19.

<sup>3</sup> Sience Daily. Web crawler. Em WWW <[https://www.sciencedaily.com/terms/web\\_crawler.htm](https://www.sciencedaily.com/terms/web_crawler.htm)>, acesso em 05.set.19.

serviço pela exaustão do servidor que não suporta a quantidade de requisições que um robô da Internet pode requisitar).<sup>4</sup>

### 3. PARTE 01 - MVP – MINIMUM VIABLE PRODUCT.

# by Wilson Amore

#### MVP

O Produto Mínimo Viável ou, no Inglês “MVP” do DANZO consistirá em disponibilizar os dados em uma página na Internet sobre o índice de violência na região da FATEC e Parque Tecnológico.

Inicialmente, os usuários serão os alunos da FATEC que, para ter acesso às informações da página terão que fazer um cadastro solicitará:

- ✓ nome, e-mail, idade, endereço;
- ✓ Qual o objetivo de ter acesso aos dados fornecidos pela página?
  - Compra/Aluguel de casa;
  - Abertura de empresas/comércios;
  - Estudo na região;
  - Trabalho na região;
  - Prática esporte na região;
  - Lazer (evento cultural, maratona geek, ou festa na região);
  - Morador/Frequentador da região.

Subsequentemente ao cadastro do usuário, a tela do usuário retornará à página inicial de login, utilizando o e-mail cadastrado. Ao entrar na página, o usuário terá acesso a informações de acordo com as opções de entrada, podendo inicialmente fazer o filtro do período (ex: de junho/2017 a agosto/2019 e tipo de violência (ex: tentativa de ou, furto/roubo de carro, moto, bicicleta, casa, pertences pessoais etc.). Essas informações serão indicadas em um Gráfico para melhor visualização.

---

<sup>4</sup> Association for Computer MAchinery. Cyber Terrorism: The New Asymmetric Threat. Em WWW <[https://www.acm.org/binaries/content/assets/public-policy/usacm/privacy-and-security/cybersecurity/03\\_07\\_24\\_spafford1.pdf](https://www.acm.org/binaries/content/assets/public-policy/usacm/privacy-and-security/cybersecurity/03_07_24_spafford1.pdf)>, acesso em 05.set.19.

Uma ideia **para o futuro** seria criar uma extensão da página (aplicativo) que dê as informações, e que também elabore uma rota mais segura para o usuário, podendo ser monitorada.

Outra ideia para captação de recursos seria tratar os dados de entrada de usuários de forma que eles se tornem atraentes para clientes potenciais interessados como:

- Empresas de segurança: com esses dados poderiam definir uma demanda de vendas/colaboradores por região;

- Imobiliárias/Construtoras: com esses dados poderiam traçar o perfil de clientes para determinadas regiões;

- Empreendedores: que teriam acesso a informações sobre o perfil da pessoa que frequenta a região, abrindo um restaurante ou loja que atenda a este público alvo.

## 4. PARTE 02 - JUSTIFICATIVAS.

### 4.1. POR QUE E COMO SERÁ UTILIZADA METODOLOGIA SCRUM?

# by Denise Oliveira

Metodologia *Scrum*:

Scrum é uma metodologia utilizada para gerenciar o trabalho em produtos complexos desde o início de 1990.

O termo “scrum” vem do meio esportivo: no jogo de rugby este termo refere-se ao reinício da partida após uma infração leve, tratando-se, em elaboração de um projeto como uma área ou tela estrutura (framework estrutural) para colacionar ideias. É composta por **ciclos** de atividades programadas — os *sprints* —, com planejamento de tarefas e datas de início e de fim determinados.

Um framework no qual pessoas podem solucionar problemas complexos e soluções adaptativas, enquanto criam de forma produtiva agregando o mais alto valor possível.

Ferramentas de *Scrum*

- *Sprints*

No Scrum, o trabalho é realizado em interações ou ciclos de até um mês, chamado de *Sprints*. Durante o planejamento do *sprint*, a equipe de desenvolvimento e o cliente (neste caso, *product owner*) devem chegar a um acordo sobre qual o **objetivo do *sprint***. Com este objetivo traçado, determinam quais os itens da lista de pendências (*backlog*) devem ser priorizados para serem executados neste *sprint*.

- Daily Scrum

Todos os dias, idealmente no mesmo horário, os membros da equipe devem realizar uma reunião (15 minutos ou menos), chamado *Daily Scrum*. Comumente nesta reunião o líder do projeto (*Scrum Master*) irá perguntar para cada membro da equipe essas três perguntas:

1. **O que fiz** ontem que ajudou o time a atingir a meta do *sprint*?
2. **O que vou fazer** hoje para ajudar o time a atingir a meta do *sprint*?
3. Existe algum **impedimento/dificuldade** que não permita a mim ou ao time atingir a meta do *sprint*?

- Kanban Scrum - Mostrar o trabalho de forma visual usando um quadro/tela:

Em geral, perde-se muito tempo em reuniões com apontamentos sobre o que está sendo feito, o que foi concluído e o que está parado. Uma maneira mais simples de resolver isso é criar um registro visual para demonstrar o curso das tarefas. Para isso, o quadro Kanban Scrum é a ferramenta ideal. Basta criar um quadro com todas as atividades (e o status de cada uma) e colocá-lo onde todos possam ver. Assim, não é preciso perder tempo falando e as pessoas já saberão sobre o status do projeto apenas observando.

- Dinâmica do Scrum

Tudo começa com a visão do produto final, que é determinado pelo *product owner*, pode ser através de um *business case*, por exemplo, ou qualquer outra visão macro do produto. Em seguida isso será desmembrado em todas as funcionalidades necessárias, que é chamada de *product backlog*. Essas funcionalidades serão organizadas por ordem de prioridade:

Assim como outros métodos com um enfoque semelhante, a proposta do *scrum* é prover contribuição para o gerenciamento de atividades intrincadas, porém de maneira flexível e que promova a adaptação do projeto diante das inevitáveis alterações. São três as ideias principais em que o método Scrum se ampara:

- Transparência
- Inspeção
- Adaptação

A noção de transparência deve ser compreendida como a existência de uma **concordância mútua entre todos os participantes do projeto**, regras que caracterizam processos e andamento dentre outros.

Outro ponto importante é a inspeção do que está sendo feito. A verificação contínua do que é produzido, seja nas reuniões diárias ou no *sprint review* é fator determinante para que se tenha a certeza de que o projeto caminha dentro do prazo estipulado, bem como para diagnosticar desvios indesejáveis e atuar de forma corretiva sobre estes últimos.

E por fim a adaptação vai de encontro a um dos principais objetivos dos métodos ágeis: a flexibilidade para mudanças. O produto que está sendo construído durante o projeto, não raro, sofre mutações. Desde que preservados os valores e práticas, pode-se adaptar os processos do *scrum* para a realidade de seu projeto/empresa.

No final do *sprint*, **duas atividades são fundamentais**: o *sprint review*, onde o objetivo é validar e adaptar o produto que está sendo construído, verificar se o que está sendo feito está de acordo com o esperado; é a apresentação do que foi feito no *sprint*. É neste ponto que surgem as mudanças e quando se atualiza o backlog.

Para finalizar, segue uma demonstração de fluxo do sprint:



## B. POR QUE E COMO SERÁ UTILIZADO O GITLAB?

# by Wilson Amore.

### GIT

Criado por Linus Torvalds em 2005 para desenvolvimento do Kernel do Linux, o Git acabou tornando-se uma ferramenta popular para os desenvolvedores de software.

O GIT é um sistema utilizado para fazer o controle de versões de arquivos, com ele é possível criar um repositório e fazer alterações no código sem o risco de sobrescrevê-los, sendo possível retornar a versão anterior do arquivo caso seja necessário.

Com o Git é possível que um grupo de pessoas trabalhe simultaneamente no mesmo arquivo, tornando o processo de desenvolver códigos mais ágil, evitando confusão de versões. Para utilizar o GIT é necessário baixar o programa pelo site <https://git-scm.com/>.

### GitLab

Desenvolvido em 2011 por Dmitry Zaporozhets e Valery Sizov, o GitLab é um software livre gerenciador de repositórios. Com ele, temos a possibilidade de criar um projeto onde várias pessoas possam contribuir no desenvolvimento, dividi-lo em tarefas e saber como está o andamento do mesmo. Para utilizar o GitLab é necessário criar uma conta no site (<https://gitlab.com/>), assim você poderá criar um novo projeto ou ser adicionado a algum existente através do seu e-mail ou usuário. Importante citar que para utilizá-lo é necessário integrá-lo ao Git, para poder enviar e receber os arquivos que ficarão armazenados no repositório online. Existem outros gerenciadores de repositórios, um dos mais conhecidos são: GitLab, GitHub, Google Code, BitBucket entre outros.

Para o desenvolvimento do nosso projeto utilizaremos o GitLab, com ele será possível dividir as tarefas do grupo e saber o andamento das mesmas, nele tem uma ferramenta similar a um “kanban”, onde nós adicionamos as atividades (*issues*) e todos os integrantes do grupo terão acesso aos arquivos, poderão verificar como está o andamento das atividades e poderão contribuir para o desenvolvimento das mesmas. As atividades concluídas terão seus arquivos anexados em seus respectivos itens assim que finalizadas. Além de todas essas funções o GitLab é gratuito, diferente do GitHub, que é pago dependendo do número de usuários e caso seja necessário utilizar repositórios privados onde o código não fica aberto para o público.

### **C. POR QUE E COMO SERÁ UTILIZADO O SLACK?**

# by Caroline Sousa

Slack

O Slack foi lançado em 2014, por um Startup desenvolvedor de Software de mesmo nome, com a principal função de se tornar um facilitador da comunicação interna na empresa, através das suas funcionalidades semelhantes a um chat, com intuito de diminuir as trocas de e-mails, reuniões e informações divergentes entre equipes.

Slack no Projeto Integrador

O Slack é uma das principais ferramentas utilizadas para o desenvolvimento do Projeto Integrador, pois é um meio de comunicação ágil e instantâneo, por este canal de comunicação, são decididos os próximos encontros presenciais, saneamentos de dúvidas das atividades desenvolvidas. Foram criados seis canais de comunicação no grupo do Slack.

- Avisos importantes: São postadas as informações mais importantes referentes ao projeto, como a data da entrega da Sprint, data da próxima reunião presencial da equipe, atividades que devem ser entregues.

- Geral: Canal destinado para discussões das atividades realizadas do projeto e dúvidas.

- Library: Este canal é apenas para compartilhamento de cursos, livros, programas e qualquer outra informação que ajude no estudo do projeto e no curso Banco de Dados.

- Outros assuntos: É um canal mais informal, podendo ser conversado sobre qualquer assunto, como compartilhamento de notícias, vagas de empregos, troca de conhecimentos, etc.

- Web bot: Neste canal serão postadas apenas informações dos avanços e entregas referentes ao projeto Web bot deste desta equipe de projeto.

- Daily: Utilizado para o envio da das informações diárias da metodologia *scrum*, que são todas as segundas, quartas e sextas até às 18h00.

## **D. POR QUE E COMO SERÁ UTILIZADO O PYTHON?**

# by Caroline Sousa

Python é uma linguagem de programação de alto nível, desenvolvida por Guido van Rossum em 1991. Os objetivos desta linguagem são desenvolver códigos bom, ágil e objetiva. Além da linguagem de fácil interpretação o Python conta com uma biblioteca que fornece várias coleções de funções.

Para o desenvolvimento do Projeto Integrador (Web Bot), foi escolhida a linguagem Python pelos integrantes do grupo por estarem familiarizados/por recomendação de especialistas e será utilizado para desenvolver as raspagens do bot.

## **5. PARTE 03 – DIAGRAMAS DAS ENTIDADES QUE SERÃO CONSTRUÍDAS.**

- a. Banco de Dados;
- b. Processos
  - i. Processo 01;
  - ii. Processo 02;
  - iii. Processo 03;
  - iv. Processo n;

## **6. CRONOGRAMA.**

O cronograma deste projeto será desenvolvido após a validação da primeira entrega datada para 06/09/19, prazo após o qual será apresentado o cronograma



geral do Projeto Integrador ditado pelos professores , a instância superior ao qual este projeto Web Bot está submetido.

## **7. CONSIDERAÇÕES FINAIS.**

Os assistentes de projeto estudaram a documentação necessária para cada tópico desta descrição de projeto, concluindo, por consenso/por maioria, de que o objeto e os processos acima tendem a ser suficientes para o desenvolvimento do projeto em sua integralidade, dentro do prazo estipulado e de acordo com o conhecimento atual e dos recursos definidos pelo Projeto Integrador da FATEC. Após, será definindo a segunda entrega e, possivelmente, as demais até data final ainda não determinanda pelos professores.