TÓPICOS AVANÇADOS EM INTELIGÊNCIA ARTIFICIAL

BACHARELADO EM SISTEMAS DE INFORMAÇÃO
UNIVERSIDADE FEDERAL RURAL DE PERNAMBUCO
UNIDADE ACADÊMICA DE SERRA TALHADA

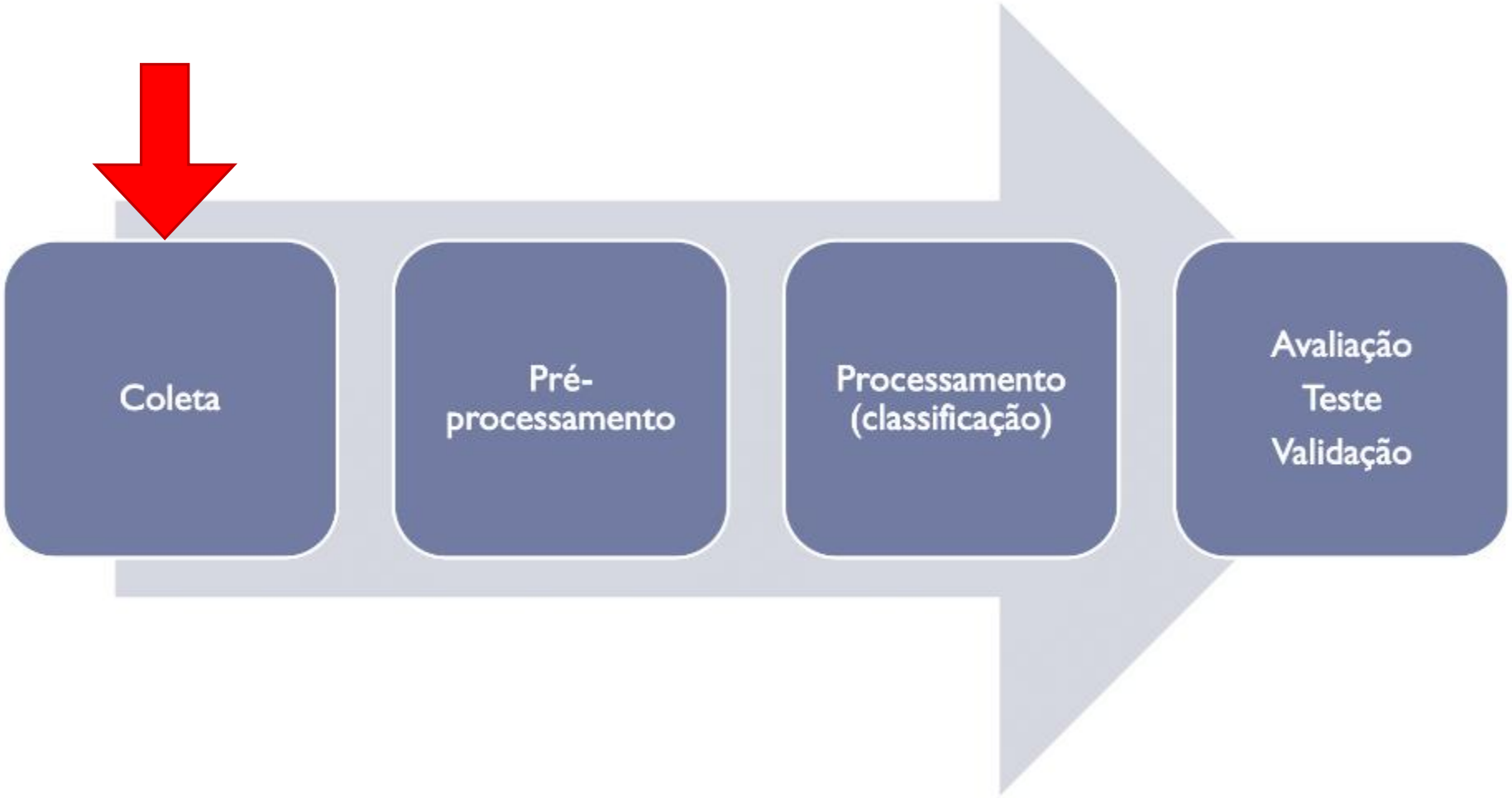
PROFESSOR: DOUGLAS VITÓRIO (douglas.alisson17@gmail.com)

COLETA DE DADOS

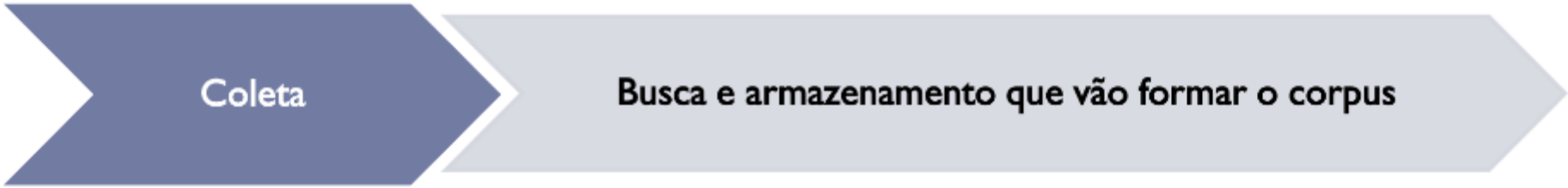
AULA 03

UFRPE

PROCESSO DE MINERAÇÃO DE OPINIÃO



CORPUS



Corpus:

- É uma coleção de **documentos;**
- Mineração de texto → documentos de texto;
- Mineração de opinião → documentos de texto contendo opiniões.

COLETA DE DADOS

A **coleta de dados** é uma das principais etapas do processo de pesquisa.

Reúne dados de interesse para análise, pesquisa, experimentação...

Técnicas comuns de coleta de dados: entrevistas, questionários (hoje de forma online).

Com a Internet veio a **automatização** da coleta de dados, Big Data...

TIPOS DE TEXTO

O texto que lidamos na **mineração de texto** pode ser:

- Formal x informal

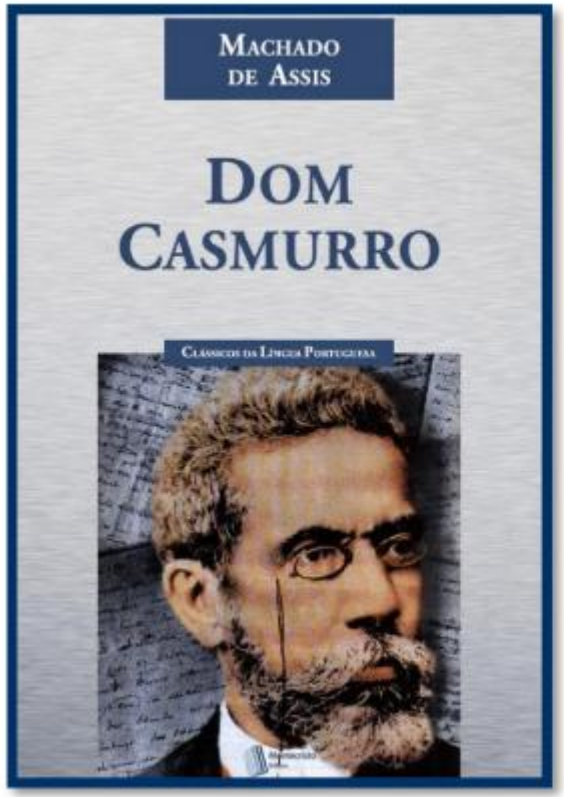
- Longo x curto

	Formal	Informal
Longo	Textos escritos em linguagem formal e que apresentam um grande tamanho (quantidade de palavras, parágrafos).	Textos escritos em linguagem informal (coloquial) com uma grande quantidade palavras e parágrafos.
Curto	Textos escritos em linguagem formal, mas com um tamanho reduzido (possuindo poucas palavras).	Textos escritos em linguagem informal (coloquial) e em uma quantidade pequena de palavras.

TIPOS DE TEXTO

1. Texto **formal e longo:**

- Artigos científicos, livros, redações...



DESENVOLVIMENTO ÁGIL: UMA ABORDAGEM SOBRE O SCRUM

Klaus Fischer Gomes Santana[1]
Rafael Araújo de Freitas[2]

RESUMO

A metodologia Scrum assume-se como extremamente ágil e flexível. Define-se como um processo de desenvolvimento iterativo e incremental que pode ser aplicado a qualquer produto ou no gerenciamento de qualquer atividade complexa, proporcionando um excelente entrosamento entre as equipes de desenvolvimento. Com todo esse entrosamento a participação ativa dos clientes, o rendimento do projeto aumenta e os requisitos e a solicitação de alteração passa a ser atendido mais rapidamente. As metodologias de desenvolvimento ágil vem se destacando a cada dia, porém essas ainda são pouco difundidas no meio acadêmico. O objetivo deste artigo, além de difundir esse assunto e servir de apoio para futuras pesquisas, é demonstrar de maneira simples e objetiva, o funcionamento, as características, o vocabulário e a aplicação da metodologia Scrum em um ambiente de trabalho.

Palavras-chave: Engenharia de Software, Métodos de Desenvolvimento, Desenvolvimento Ágil, SCRUM.

1 INTRODUÇÃO

Um desafio constante da área de Engenharia de Software é melhorar o processo de desenvolvimento de software. Mesmo com a constante evolução de métodos, técnicas e ferramentas, a entrega de softwares em prazos e custos estabelecidos nem sempre é efetiva. Uma das causas desse problema é o excesso de formalidade nos modelos de processo propostos nos últimos 30 anos.

Existe hoje a necessidade de criar software de forma mais rápida, mas com qualidade. Esse produto pode ser obtido utilizando métodos ágeis e padrões organizacionais e de processos.

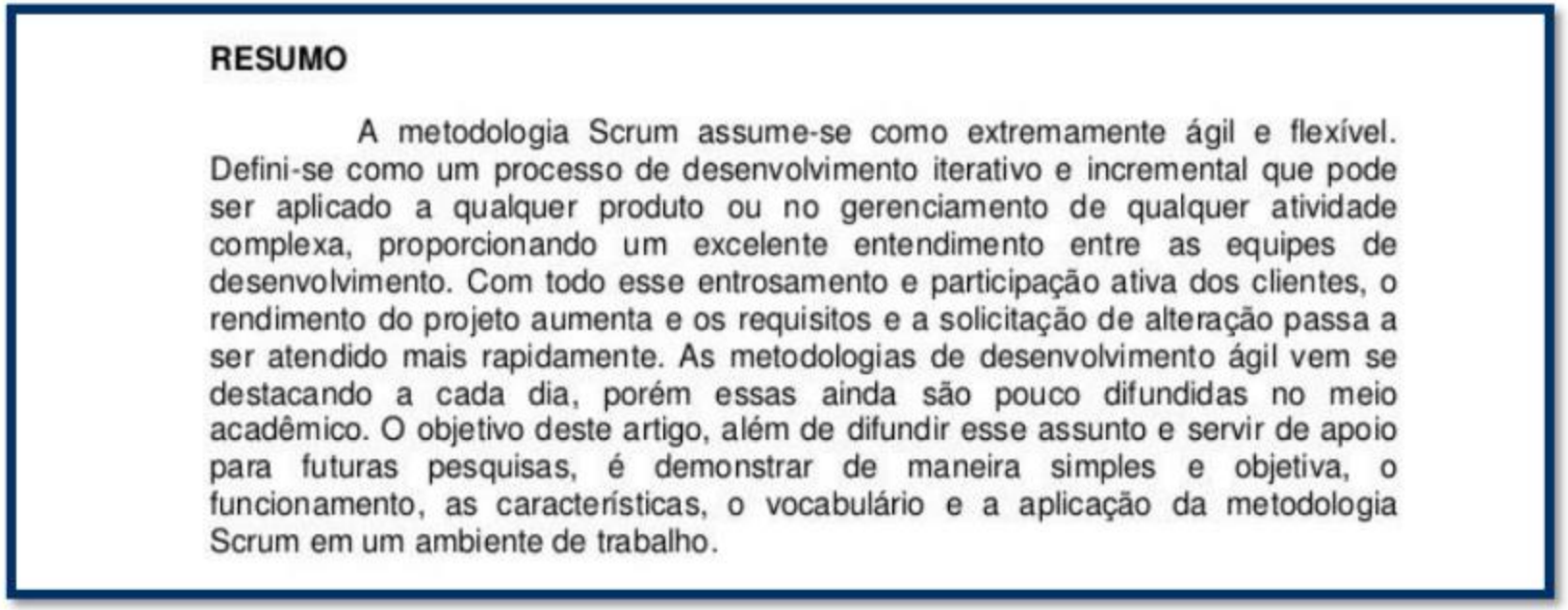
O desenvolvimento ágil é uma filosofia. É uma maneira de pensar sobre produção de software. A decisão canônica desta maneira de pensar é o "Manifesto Ágil" (Beckel al., 2001), um conjunto de 4 valores (figura 1) e 12 princípios (figura 2).

[1] Graduando em Análise e Desenvolvimento de Sistemas. E-mail: klausfgsantana@gmail.com
[2] Graduando em Análise e Desenvolvimento de Sistemas. E-mail: jedi.rafael@gmail.com

TIPOS DE TEXTO

2. Texto **formal e curto:**

- Abstracts (resumos de artigos), parágrafos...



RESUMO

A metodologia Scrum assume-se como extremamente ágil e flexível. Defini-se como um processo de desenvolvimento iterativo e incremental que pode ser aplicado a qualquer produto ou no gerenciamento de qualquer atividade complexa, proporcionando um excelente entendimento entre as equipes de desenvolvimento. Com todo esse entrosamento e participação ativa dos clientes, o rendimento do projeto aumenta e os requisitos e a solicitação de alteração passa a ser atendido mais rapidamente. As metodologias de desenvolvimento ágil vem se destacando a cada dia, porém essas ainda são pouco difundidas no meio acadêmico. O objetivo deste artigo, além de difundir esse assunto e servir de apoio para futuras pesquisas, é demonstrar de maneira simples e objetiva, o funcionamento, as características, o vocabulário e a aplicação da metodologia Scrum em um ambiente de trabalho.

TIPOS DE TEXTO

3. Texto **informal e longo:**

- Textos de blogs pessoais, e-mails, análises de filmes/livros/locais...

★★☆☆☆ minha estante

William 18/03/2012

O Pequeno Principe
Sabe aquela sensação que você tem quando percebe que não deu o valor necessário a alguma coisa? Antes de começar essa resenha tenho que dizer que isso aconteceu com essa obra. A um termo muito famoso em games que descreve o que achei desse livro na primeiar vez: "Overused", seria algo como "muito adorado mas é ruim" para traduzir melhor, há coisas que são Overuseds, Jhonny Depp por exemplo, TODO personagem dele é a mesma coisa, um cara louco mas com carisma. Isso não é um problema já que ele atua bem como esse personagem, o problema é que ele SEMPRE faz a mesma atuação, em "Edward Mãos de Tesoura", "A Fantastica Fabrica de Chocolates", "Alice no País das Maravilhas", "Piratas do Caribe"... E mesmo assim as pessoas ainda acham ele um "ator versátil".

Quando li "O Pequeno Principe" pela primeira vez, com meus 7/8 anos, ouvi tantas maravilhas sobre o livro que fui esperando encontrar algo maravilhoso, e na época não encontrei, por isso o coloque inessa lista pessoal de "Overused", até que tive que que relê-lo e aí sim pude perceber uma coisa: Como ele é profundo e adulto, a primeira vista a história simples pode até parecer infantil, mas se analisada profundamente (como pude fazer agora e não podia na primeira vez) se mostra uma das melhores obras da literatura mundial.

Talvez pelo fato do principe ter algo que encanta todo tipo de publico: A inocência característica das crianças, A paixão pela flor mesmo com espinhos, típica da adolescencia, e a vontade de descobrir a razão da vida dos adultos, ele se mostra um dos poucos livros que não é indicado apenas para uma unica faixa etária, e pode cativar tanto uma criança como um idoso em suas poucas páginas, contando a história do pequeno habitante do planeta B612 através do seu processo de crescimento.

A história começa pela visão de um aviador que está perdido no deserto e vivia uma vida simples pois deixou seu talento para desenho ser cercado pela monotomia dos adultos, mas que ao encontrar o principe vê a beleza nas coisas simples, e logo passa para o Principe, que após morar num planeta pequeno decidiu sair para compreender as complexas personalidades que podem ser encontradas no mundo. Junto com ele passamos pelo Rei, que é viciado em mandar mas nos ensina que só podemos exigir o que as pessoas podem nos dar, o Vaidoso, que nos mostra que não necessitamos da atenção dos outros mas precisamos reconhecer nossos próprios valores, o Bebado que tenta escapar de uma realidade que ele mesmo cria, o Homem de Negócios que nos mostra que não podemos gastar todo nosso tempo com as coisas importantes da vida, o Geólogo que sabe das coisas nas teoria mas não se arrisca a sair para ver se encontra algo novo, entre outros personagens que tem todos uma coisa em comum: Todos mostram traços que sempre estarão presentes na humanidade, e o principal, sem tomar partido em nenhum deles, o autor apresenta a situação pelo olhar ingênuo do principe e deixa que o leitor tome suas próprias conclusões sobre a situação.

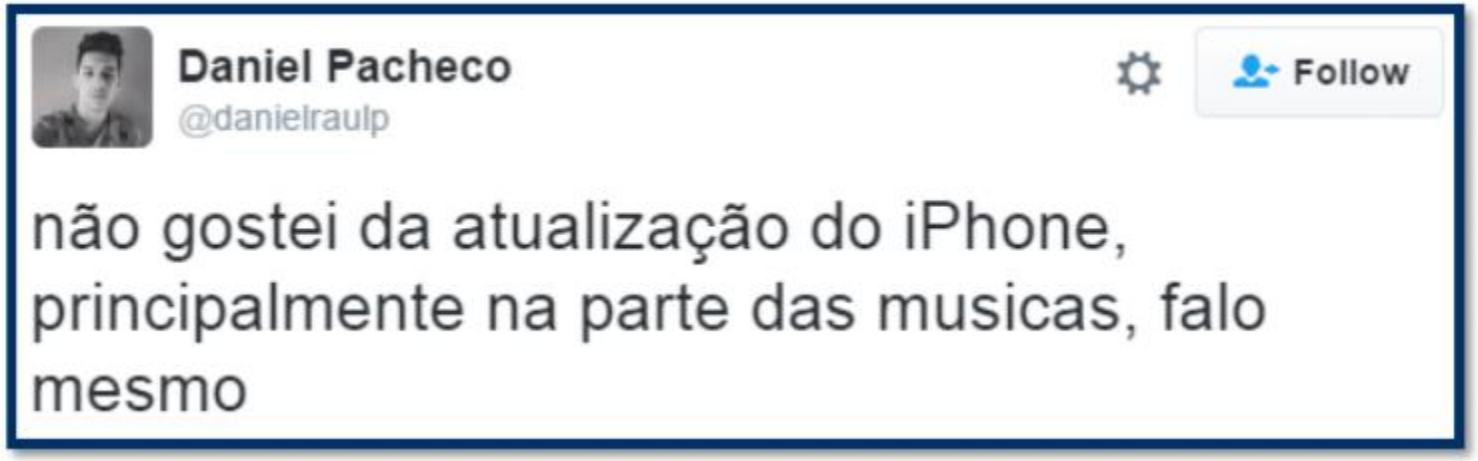
No final, "O Pequeno Principe" se mostra um livro diferente, não pelo fato de poder ser apreciado por todos, ou por ser a história mais contada de todos os tempos, mas sim por ser um livro que não apenas se lê, mas que também se entende, algo muito raro de se ver...

gostei (25) comentários(0) comente

TIPOS DE TEXTO

4. Texto **informal e curto:**

- Comentários de Facebook, *tweets*, análises de filmes/livros/locais...



BALANCEAMENTO DAS CLASSES

Corpus balanceado:

- Todas as classes possuem o mesmo número de documentos

	Classe A	Classe B	Classe C	Total
Corpus 1	300	300	300	900

Corpus desbalanceado:

- Cada classe possui uma quantidade diferente de documentos

	Classe A	Classe B	Classe C	Total
Corpus 1	300	450	150	900

BALANCEAMENTO DAS CLASSES

O ideal é que não tenhamos corpora desbalanceados.

Então, para isso, reduzimos a quantidade de dados na base...

	Classe A	Classe B	Classe C	Total
Corpus 1	300	450	150	900

	Classe A	Classe B	Classe C	Total
Corpus 1	150	150	150	450

COLEÇÃO DOURADA

Conhecida também como *Golden Standard* ou *Golden Collection*.

Consiste no conjunto de **dados rotulados** para treinamento na **Aprendizagem Supervisionada**.

É um **padrão** de referência para a classificação.

Geralmente são construídas **manualmente**, por isso demandam **muito tempo e esforço**.

COLEÇÃO DOURADA

E esse shopping cheio de casalzinho? Eca	Negativo
Mais um dia dando uma de fotógrafo, mano como eu amo isso	Positivo
Fui no pdb e já peguei a roupinha de hj né	Neutro
Vou carregar meu cel um pouco	Neutro
Tenho que distrair a cabeça, ficar trancado dentro de casa não ajuda	Negativo
Ir pra acad nesse frio é mó desanimo	Negativo
Fico tão fofo de barba.	Positivo
Vou fazer minha unha	
@suhoxbae parece ser realmente bom	
Só sufoco mas a noite vai fazer valer a pena	

DADOS PARA MINERAÇÃO DE OPINIÃO

- **Sites de comércio e serviço eletrônico:** amazon.com, booking.com...
- **Mídias online:** jornais, blogs...
- **Redes sociais:** UGC, Twitter, Facebook...

CORPORA DISPONÍVEIS (BENCHMARK)

IMDb Movie Review data (Pang e Lee):

- Avaliações de filmes em inglês
- http://www.cs.cornell.edu/people/pabo/movie-review-data

Stanford Sentiment140:

- *Tweets* com opiniões de consumidores, eleitores, fãs, etc. em inglês
- http://help.sentiment140.com/for-students/

Sanders:

- *Tweets* acerca de empresas (Apple, Microsoft, Google, Twitter) em inglês
- https://github.com/zfz/twitter_corpus

COMO COLETAR DADOS?

Ferramentas:

- *Social media data collection tools*: http://socialmediadata.wikidot.com

Programação (via API):

- Interface de Programação de Aplicativos (API)
- REST, Streaming (Twitter)
- Graph API (Facebook)

Programação (Web Scraping):

- Raspagem de dados
- Extração de dados de páginas da Web

COLETANDO DADOS DO TWITTER

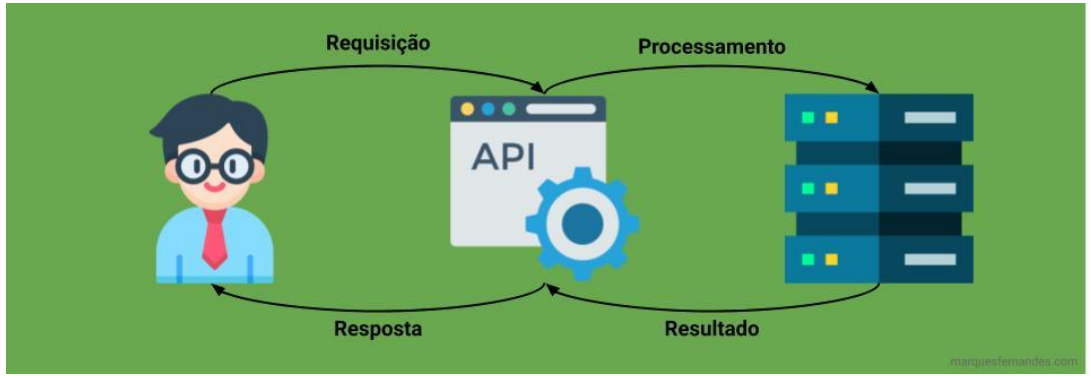
API

Uma **API (*Application Program Interface*)** é uma **interface de programação de aplicativos.**

Com ela, produtos e serviços são acessados **externamente.**

Mapas, pagamentos, dados...

Traz simplicidade para os desenvolvedores.



APIs DO TWITTER

https://developer.twitter.com/en/docs/twitter-api

REST API e **Streaming API**.

Request-Response (Requisição-Resposta).

Requisições são autenticadas pelo protocolo **OAuth**:

- Padrão aberto de autenticação para login em sites ou redes sociais de forma segura.
- Google, Facebook e Microsoft também utilizam o OAuth.

REST API

A **REST (*Representational State Transfer*) API** é usada para **recuperar *tweets* antigos.**

A aplicação envia uma requisição ao servidor e o servidor se comunica de volta com o cliente com a resposta e a comunicação é **encerrada.**

Adota o formato JSON como padrão de resposta.

As requisições possuem limites: dentro de uma janela de 15 minutos, um total de 180 (por usuário) ou 450 (por aplicação) *tweets* podem ser coletados.

STREAMING API

A **Streaming API** é usada para **recuperar *tweets* recentes,** na hora em que eles são postados.

Mantém a conexão HTTP aberta para continuar recebendo respostas sempre que existirem atualizações.

Indicada para aplicações que irão processar ou monitorar dados em **tempo real**.

Adota o formato JSON como padrão de resposta.

Possui taxas de limites, porém não há valores exatos (+/- 50 mil/dia).

BIBLIOTECA TWEEPY

Biblioteca Python para acessar a API do Twitter.

Instalação:

pip install tweepy

GitHub:

git clone https://github.com/tweepy/tweepy

cd tweepy

python setup.py install