

## 1 Introdução

A extração de dados *on-line* tornou-se comum nos últimos anos, dados são extraídos da *internet*, armazenados e trabalhados, gerando algum tipo de abstração - significado final. As coletas de dados da *internet* podem ser feitas por vias manuais ou automatizadas. As desvantagens da extração manual são o tempo demandado e a possibilidade de erro humano durante o processo, ainda mais a extração de uma grande quantidade de dados. Já a extração automatizada possui as vantagens da velocidade e da precisão, por se tratar de um dispositivo de computador que efetua tarefas a partir de instruções ministradas por meio de um código.

A retirada em massa de dados de *sites* para possibilitar uma análise é conhecida por diversas nomenclaturas: coleta de dados *web*, raspagem, extração massiva, mineração e *web scraping*. Existem alguns programas que oferecem os recursos de *Web Scraping*, como: *Octoparse*, *Mozenda*, *Import.io*, *Content Grabber*, dentre outros. Esses programas oferecem algumas dificuldades:

- Não oferecem ferramentas de extração para sites com arquitetura *HTML* distinta.
- Alto preço para utilização.
- Requerem do usuário conhecimentos computacionais.
- Alguns sites possuem bloqueio contra programas de extração massiva de dados (uma prevenção para um extrator mal-intencionado).

plataformas digitais de banco de dados de patentes possuem grande quantidade de dados, plataformas de arquiteturas estruturadas e não estruturadas, bem como proteções contra furto. Destarte, a captação de dados de patentes nessas plataformas é inviável a forma manual, bem como por ferramentas de *Web Scraping online* já conhecidas. E para a resolução do problema de extrair e analisar/estudar dados de patentes dos bancos de dados digitais dos diversos escritórios mundiais, é necessário:

- a. Utilizar um *software* específico para cada plataforma, e que consiga navegar por arquiteturas estruturadas ou não;
- b. Que possua uma interface simples ao usuário e que não exija conhecimentos avançados em programação;
- c. Com um comportamento semelhante à ação humana de busca, de maneira que não seja bloqueado pelos mecanismos de defesa de “furto” de dados.

Mediante as dificuldades de obtenção de tais dados, surge a necessidade da construção de *softwares* que atendam ao requisito de captar dados de patentes *on-line*, de forma estruturada e a possibilidade de se estudar os dados com alguma ferramenta de *Business Intelligence* - *BI*. Foram escolhidas três plataformas de patentes para a construção de *softwares*: *Derwent – Web of Science*, *Espacenet* e *US Patent*. As escolhas dessas tiveram como critério o fato delas serem abrangentes, o seja, possuírem cobertura de diferentes regiões, e as duas últimas serem gratuitas. A escolha da *Derwent – Web of Science* ocorreu pela sua abrangência continental e também pela possibilidade de acesso da instituição de ensino da pesquisa.

Objetivos propostos a serem alcançados:

- Construção de um aplicativo de *software* para a plataforma do *Web of Science*;

- Construção de um aplicativo de *software* para a plataforma *Espacenet*;
- Construção de um aplicativo de *software* para a plataforma *US Patent*;

Este trabalho contém cinco tópicos: o primeiro uma contextualização, seguido dos objetivos; o segundo apresenta o referencial teórico; o terceiro demonstra a metodologia de pesquisa utilizada; o quarto a análise dos resultados obtidos; e, por último trata-se das conclusões e das considerações finais.

## 2 Referencial Teórico

O referencial teórico apresentado demonstra os materiais/conteúdos utilizados para embasar o desenvolvimento do trabalho: bases digitais de patentes, inteligência artificial, linguagem de programação Python, e por fim qualidade e avaliação de *software*.

### 2.1 Bases digitais de dados de patentes

Os registros de patentes pelo mundo estão em diferentes agências e estão disponibilizados em diversas plataformas. Rodrigues (2015) cita a existência de algumas bases de dados de patentes digitais utilizadas.

Quadro 1- Plataformas digitais de patentes.

BASE	Endereço eletrônico
Instituto Nacional da Propriedade Industrial	<a href="http://www.gov.br/inpi/pt-br">www.gov.br/inpi/pt-br</a>
<i>Google Patents</i>	<a href="http://patents.google.com/">patents.google.com/</a>
<i>Espacenet – European Patent Office</i>	<a href="http://worldwide.espacenet.com/">worldwide.espacenet.com/</a>
<i>US Patent – Patent and Trademark Office</i>	<a href="http://www.uspto.gov/">www.uspto.gov/</a>

Fonte: Adaptado de Rodrigues, 2015.

Fiocruz (2021) também cita algumas outras bases.

Quadro 2- Plataformas complementares de bases de patentes.

BASE	Endereço eletrônico
<i>Web of Science – Clarative Analytics</i>	<a href="http://login.webofknowledge.com/">login.webofknowledge.com/</a>
<i>European- Patent Office – EPO</i>	<a href="http://european-patent-office.org">european-patent-office.org</a>
<i>Patents – IBM</i>	<a href="http://delpion.com">delpion.com</a>
<i>Italian Patent and Trademark Office</i>	<a href="http://european-patent-office.org">european-patent-office.org</a>
Instituto Nacional da Propriedade Industrial de Portugal	<a href="http://inpi.pt">inpi.pt</a>

Fonte: Adaptado de [www.castelo.fiocruz.br/vpplr/banco\\_patentes.php](http://www.castelo.fiocruz.br/vpplr/banco_patentes.php).

Para Falagas *et al.* (2008), as diversas bases de dados referentes às patentes não abrangem toda a área da ciência igualmente, ou seja, cada uma das bases possui suas próprias peculiaridades.

### 2.2 Inteligência artificial

Monard e Baranauskas (2000) relata que a busca da sociedade por preceitos ou maquinismos aptos a fazer o serviço humano, bem como a imitar o raciocínio do ser humano vem sendo objeto de estudo. O autor cita que o conceito de dar ao computador ou ao

microcomputador a aptidão de desempenhar funções que o ser humano conseguiria resolver com a utilização do cérebro vem se impregnando em alas da ciência.

Monard e Baranauskas (2000) descrevem a IA como uma ala da Ciência que tem o intuito de tornar aptos os dispositivos de computadores a pensarem ou se comportarem de forma inteligente executando algoritmos e tratando informações de forma ordenada. Os autores citam que a IA é ampla e abrangente, estando também relacionada com ramos como a psicologia, a biologia, a lógica a matemática, a linguística, a engenharia, a filosofia, entre outras áreas do conhecimento.

Figura 1- Áreas parciais em que a IA está contida



Fonte: Adaptado de Monard e Baranauskas (2000).

Para que os dispositivos de computadores pensem ou se comportem de forma inteligente via algoritmos tratando informações de forma ordenada, carece-se de uma linguagem computacional robusta. A linguagem de programação *Python* possui muitas funcionalidades para o desenvolvimento de algoritmos de IA.

### 2.3 Python

Menezes (2014) cita que a linguagem Python é livre, disponível para todos os computador e sistemas. Cita também que por ser linguagem de código aberto e gratuito chama atenção dos desenvolvedores. Python é uma linguagem modular, pois trabalhos já foram desenvolvidos por outros desenvolvedores e disponibilizados na comunidade digital localizada no endereço eletrônico oficial da linguagem de programação. Uma das maiores facilidades da linguagem de programação Python é que ela possui uma quantidade imensa de bibliotecas que podem ser importadas para o programa a qualquer instante.

Menezes (2014) afirma que a linguagem Python embora simplória, é poderosa. Trata-se de uma linguagem clara e objetiva. Labaki (2003) define a linguagem de programação Python como modular, explicitando que algum programa que necessita ser construído parcialmente ou totalmente pode ser encontrado em formato modular na comunidade.

A linguagem de programação Python possui um módulo/pacote chamado *Scrapy*, disponível na "*PIP*" (*Python Package Index*), um sistema de gerenciamento de pacotes padrão

usado efetuar a instalação de módulos de software em Python. Por ser um *Framework* (abstração de união de códigos), o *Scrapy* oferece várias possibilidades de trabalho (funcionalidades) que facilitam o processo de navegação na rede de forma automatizada, são eles (Scrapy, 2021):

- Controle de navegação na *web*;
- Bibliotecas de compilação em *HTML*;
- Filtragem e tratamento de dado.

A instalação do módulo *Scrapy* ocorrer pelo comando *PIP*: \$ *pip install scrapy*.

#### 2.4 Qualidade e avaliação de *software* (ISO 9126 e ISO 14598)

A norma ISO 9126 é a antiga NBR 13596, e fornece a padronização do método avaliativo de qualidade de *softwares*.

A norma ISO 9126 cita que elaborar/arquitetar produtos de *software* de alta qualidade é de basilar importância, pois a especificação e a avaliação da qualidade do produto de *software* são pontos importantes para garantir qualidade adequada. Assim, define um modelo de qualidade como um conjunto de características e o relacionamento entre elas. São essas características que fornecem a base para a especificação dos requisitos de qualidade e para a avaliação de qualidade.

A norma ISO 9126 direciona que é de suma importância que cada uma das características relevantes de qualidade do *software* seja especificada e assim avaliada fazendo uso, quando possível, de métricas. Cita também que dentre as partes que podem avaliar um *software*, um desenvolvedor pode efetuar a avaliação de um *software*, utilizando assim as medidas de características de qualidade.

A norma ISO 9126 demonstra em seu escopo um conglomerado de atributos caracterizantes de um produto de *software* com o intuito de descrever quais são os tópicos importantes a considerar na fase de avaliação de um *software*/aplicativo. O modelo de qualidade descrito na norma é subdividido em duas partes: **a)** qualidade interna e qualidade externa e **b)** qualidade em uso.

**a)** As qualidades externa e interna podem ser descritas como os atributos de qualidade de *software* em seis características (funcionalidade, confiabilidade, usabilidade, eficiência, manutenibilidade e portabilidade):

- **Funcionalidade** - aspectos de *software* como: interoperabilidade, acurácia no funcionamento, segurança de utilização/acesso e adequação;
- **Confiabilidade** - aspecto de maturidade do programa, que resulta em alta recuperabilidade e baixa gama de falhas;
- **Usabilidade** – se resume em operacionalidade e atratividade para utilização;
- **Eficiência** – comportamento em função do tempo, é eficiente desde que atenda as demandas primárias e que seus recursos sejam utilizáveis;
- **Manutenibilidade** – estabilidade, facilidade de análise e facilidade de efetuação de testes, bem como de modificação;
- **Portabilidade** - capacidade de fácil instalação, coexistência e adaptação em diferentes sistemas.

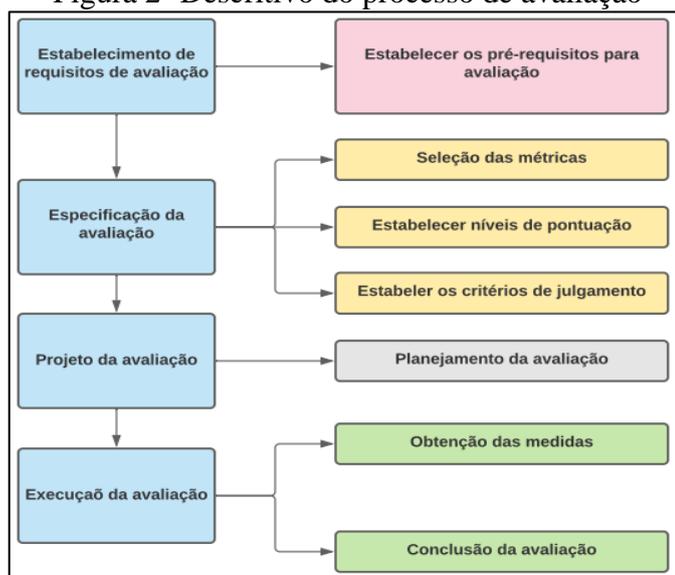
**b)** As métricas da qualidade em uso são categorizadas em quatro características (eficácia,

produtividade, segurança e satisfação.

- **Eficácia** - capacidade do produto de atingir as metas prévias especificadas;
- **Produtividade** - relação entre o emprego de recursos disponíveis e a eficácia;
- **Segurança** - capacidade do produto de *software* de apresentar níveis aceitáveis de riscos de danos a pessoas, a negócios, a *softwares*, propriedades ou ao ambiente, em um contexto de uso especificado;
- **Satisfação** - capacidade do produto de *software* de satisfazer usuários, em um contexto de uso especificado.

A norma ISO/IEC 14598 faz uma ponte com a norma ISO 9126, apresentando requisitos para a avaliação da qualidade de um programa de *software* e que é conveniente que todos os requisitos a serem alcançados sejam claramente especificados. A norma cita também que para a realização da efetuação de avaliação da qualidade de *software*, convém que a qualidade seja medida se possível usando métricas. A norma ISO/IEC 14598 define a palavra “métrica” internamente como uma escala quantitativa usada para a medição.

Figura 2- Descritivo do processo de avaliação



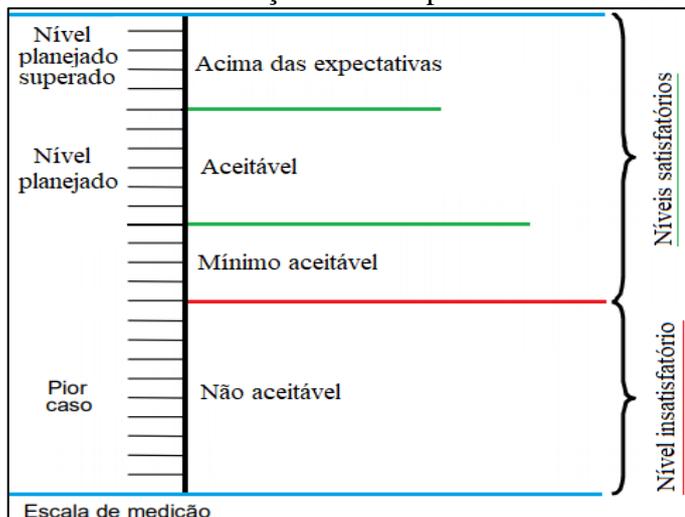
Fonte: Adaptado de ISO/IEC 14598.

O desenvolvedor, segundo a ISO/IEC 14598, é aquele que efetua a execução do desenvolvimento do *software*, incluindo os pré-requisitos, o projeto e os testes de aceitação do *software*. Essa norma descreve um fluxograma de avaliação tendo como ponto de partida o estabelecimento de pré-requisitos.

A ISO/IEC 14598 especifica que os pré-requisitos de avaliação são criados e definidos a partir dos tópicos de qualidade citados na norma ISO/IEC 9126, com o objetivo de assegurar que o produto forneça as qualidades pretendidas, atendendo assim as necessidades do usuário.

A norma ISO/IEC 14598 cita que é importante que as medições de um *software* sejam executadas de uma forma simples/ágil e define métrica como os parâmetros dos atributos que podem ser quantificáveis. Diz também que é conveniente que as métricas usadas sejam claras (objetivas), testáveis (empíricas) e utilizem uma escala válida e que possa ser reproduzida. As definições da norma são:

Figura 3- Escala de satisfação descrita pela norma ISO/IEC 14598



Fonte: Adaptado de ISO/IEC 14598.

O valor medido dentro de uma escala válida e mensurável não traduz em si um nível de satisfação quanto aos requisitos. Assim sendo, uma divisão de satisfação em formato de escala pode ser efetuada.

### 3 Metodologia

A pesquisa teve como metodologia a *Design Science Research*. A metodologia *Design Science Research* possui a proposta de construção de um artefato, palpável ou artificial. Para Bayazit (2004) o método *Design Science Research* possui um procedimento rigoroso que visa a criação de algo para resolver um problema, tendo como passo posterior à criação, a avaliação do projeto e dos seus resultados, fora a divulgação da pesquisa e do resultado. Hevner *et al.* (2004) fornecem uma série de orientações (protocolos a serem seguidos) que podem ser seguidas para a pesquisa científica.

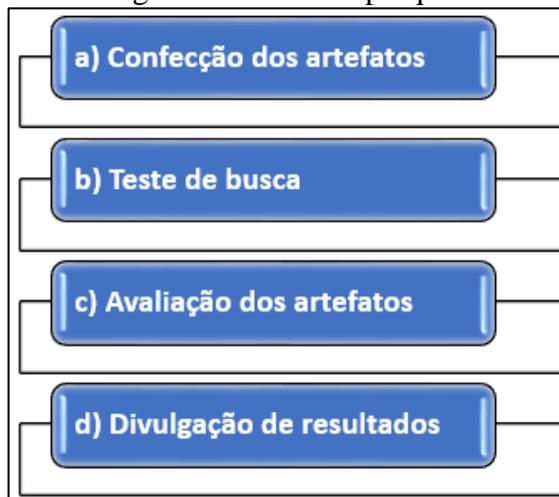
Quadro 3 - Protocolo *Design Science Research*

Orientação	Descrição
<b>Orientação 1:</b> <i>Design</i> como um artefato	Deve produzir um artefato viável em forma de uma construção, um modelo, um método ou uma instanciação.
<b>Orientação 2:</b> Relevância do problema	Desenvolver soluções baseadas em tecnologia para problemas relevantes para empresas e instituições.
<b>Orientação 3:</b> Avaliação do <i>design</i>	A utilidade, qualidade e eficácia do artefato deve ser demonstrada pelos métodos de avaliação.
<b>Orientação 4:</b> Contribuições da pesquisa	Deve fornecer contribuições claras e verificáveis nas áreas do artefato.
<b>Orientação 5:</b> Rigor de pesquisa	É baseada na aplicação de métodos rigorosos na construção e na avaliação.
<b>Orientação 6:</b> <i>Design</i> - processo de busca	Requer a utilização de meios disponíveis para alcançar os objetivos.
<b>Orientação 7:</b> Comunicação de pesquisa	Deve ser apresentada ao público voltado à área de aplicação do artefato.

Fonte: Adaptado de Hevner *et al.* (2004).

Baseado nas orientações do Quadro 3, bolou-se um passo a passo a ser seguido:

Figura 4- Passos da pesquisa



Fonte: os autores.

### 3.1 Construção dos programas mineradores automatizados – robôs de busca

A rede de *internet* se tornou uma ferramenta gigantesca para a disseminação de dados. Desta forma, ferramentas que proporcionam captação e análise de dados de forma automatizada também ganharam espaço e se tornaram necessárias.

Para se caminhar entre páginas em um navegador automatizado de uma forma não sequencial via *software*, o programa deve conter blocos ordenados de navegação. Se a navegação automatizada se der de uma forma sequencial, o programa de navegação pode conter instruções de repetição acopladas ou não às condições (*for*, *while* etc.).

As páginas virtuais possuem um componente que permite a estruturação do endereço eletrônico, bem como a organização dos acessos de caminhos das páginas e os devidos endereçamentos. Este componente é conhecido como *HTML*, que significa *Hiper Text Markup Language*. Dentre os hipertextos conectados em uma página eletrônica, podem ser utilizados: imagens, vídeos, palavras, documentos, arquivos em geral etc. Esta forma de *hiperlink* forma um caminho e a estruturação visual do endereço, bem como a sua aparência final.

Um dos módulos de Python mais utilizados para navegação automatizada é o *Selenium Webdriver*, que efetua o trabalho de interação com os *browsers* simulando a ação humana. A fim de se retirar o conteúdo de um hipertexto, a linguagem de programação utiliza a localização do *Xpath* (seleção de nós de um documento para a atribuição de valores a um documento) no código juntamente com a automatização do navegador, trazendo ao programa uma simulação da realidade em que o usuário estaria navegando pelo navegador e extraindo os dados manualmente.

Os conteúdos extraídos dos hipertextos dentro do navegador podem ser armazenados dentro de listas e transformados em uma matriz de linhas e colunas a partir de módulos matemáticos específicos. Um módulo que possibilita facilmente este trabalho é o módulo *numpy*. Os dados passados por matriciamento, a partir de módulos, podem ser armazenados e salvos em um formato de planilha eletrônica. Um módulo que efetua este trabalho com é o pacote *xlwt*.

Os *softwares* foram projetados e construídos no ambiente de desenvolvimento integrado do *PyCharm*® Educacional 2020, por se tratar de um ambiente interativo, colorido, com ferramentas completas, capaz de fornecer apontamentos diversos de ajuda e por ser uma versão gratuita.

### 3.2 Construção do *software* buscador - *WoS*

Para a construção do buscador foi necessário um entendimento sobre a plataforma do *WoS*. A partir da plataforma do *WoS* é possível analisar indicadores de produtividade científica e tecnológica. Sua base de dados de patentes é ampla e possui informações relevantes de escritórios conveniados espalhados ao redor do mundo. A base de dados possui informações relevantes de uma patente, como: inventores, informações das invenções registradas, instituições ou autores que registraram, datas importantes, áreas de utilização/aplicação, países de origem etc.

O do código foi montado em diferentes blocos para facilitar seu manuseio, foram eles:

- 1º bloco:** importação de bibliotecas;
- 2º bloco:** seleção de *driver* utilizado e abertura do navegador;
- 3º bloco:** inserção do endereço eletrônico no navegador;
- 4º bloco:** definição/seleção de idioma;
- 5º bloco:** informações de entrada e acesso à plataforma;
- 6º bloco:** definição da base de dados;
- 7º bloco:** pesquisa com o algoritmo de busca;
- 8º bloco:** abrir o primeiro resultado das pesquisas;
- 9º bloco:** sub-rotina de captação de dados das páginas;
- 10º bloco:** armazenamento dos dados.

### 3.3 Construção do *software* buscador - *Espacenet*

*Espacenet* é uma plataforma virtual grátis, que tem como foco a busca de patentes e de pedidos de patentes existentes em seu banco de dados. A plataforma foi elaborada por um Escritório Europeu de Patentes contendo os dados dos escritórios dos estados que fazem parte da Organização Europeia de Patentes. A plataforma é disponibilizada de forma gratuita na rede e é encontrada pelo seu respectivo endereço eletrônico. Ela oferece na ala de pesquisas avançadas opções de buscas refinadas em diferentes vertentes como: palavras chaves contidas em título, número de patentes, cessionários, inventores, numeração identificadora de aplicações, dentre outras.

O do código foi montado em diferentes blocos para facilitar seu manuseio, foram eles:

- 1º bloco:** importação de bibliotecas;
- 2º bloco:** seleção de *driver* utilizado e abertura do navegador;
- 3º bloco:** inserção do endereço eletrônico no navegador;
- 4º bloco:** inserção da busca;
- 5º bloco:** sub-rotina de seleção de patentes e *download*;

### 3.4 Construção do *software* buscador – *US Patent*

O Escritório de Patentes e Marcas dos Estados Unidos é uma entidade do governo federal americano que trata do fornecimento de concessão de patentes e pedidos e de registro de marcas no território dos Estados Unidos. O escritório, como um ente federal, tem como intuito o cumprimento da Constituição Americana, que cita a promoção e o progresso da ciência e das artes úteis com a garantia de utilização fornecida aos criadores por tempo delimitado.

Vinculado à Secretaria de Comércio, o escritório possui uma plataforma virtual encontrada por um endereço eletrônico que possui um banco de dados com todas as patentes e com todos os pedidos de concessão registrados em seu portfólio.

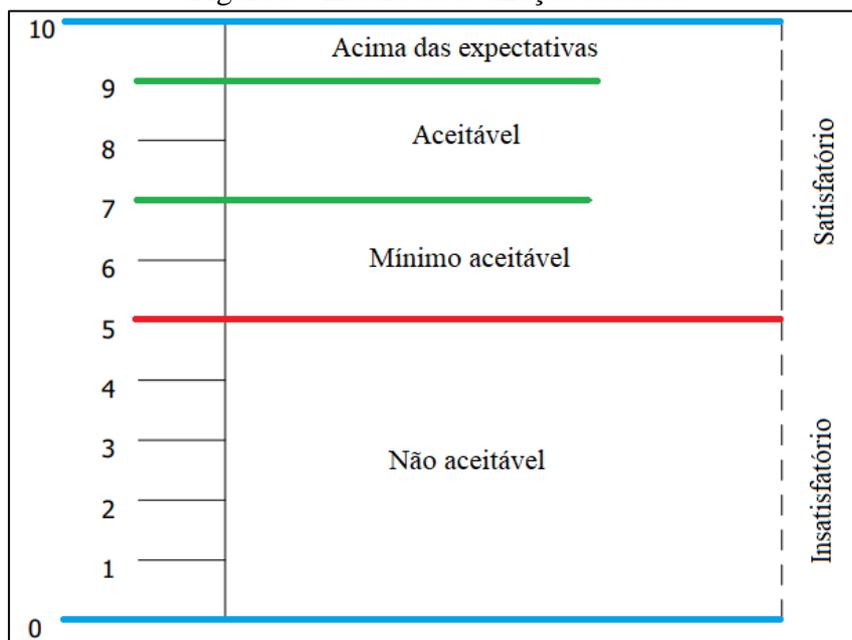
O do código foi montado em diferentes blocos para facilitar seu manuseio, foram eles:

- 1º bloco:** importação de bibliotecas;
- 2º bloco:** seleção do *driver* e abertura do navegador;
- 3º bloco:** inserção do endereço eletrônico;
- 4º bloco:** inserção do algoritmo de busca e efetuação da pesquisa;
- 5º bloco:** abertura de resultado;
- 6º bloco:** captação de dados via sub-rotina;
- 7º bloco:** armazenamento.

### 3.5 Método de avaliação dos *softwares*

Os produtos de *software* carecem de uma avaliação para classificar sua operação. Seguindo as diretrizes da norma ISO 14598, foi adotada a escala de avaliação de satisfação para a avaliação de qualidades de um *software* descritas da norma ISO 9126. A fim de medir a eficácia do funcionamento dos aplicativos no que diz respeito aos seus atributos que podem ser medidos quantitativamente, a escala de 0 a 10 foi adotada (a norma não submete o avaliador a um critério de escala específico), submetendo assim cada atributo de cada *software* a dez testes. Para os atributos que não podem ser medidos quantitativamente, ou que não carecem de medição quantitativa (a norma deixa a critério discricionário do avaliador), avaliações qualitativas são adotadas (sempre fundamentadas e explicadas). A Figura 5 demonstra a escala adotada para a avaliação dos produtos seguindo a norma ISO/IEC 14598 confeccionada a partir da Figura 3.

Figura 5 - Escala de avaliação adotada



Fonte: os autores.

Como método de avaliação dos *softwares*, foram feitas dez simulações de atuação/funcionamento com cada código, simulações feitas com algoritmos de busca diferentes,

a fim de testar o funcionamento dos programas com poucos dados encontrados e muitos dados encontrados.

#### 4 Análise dos resultados

Os tópicos a seguir têm o intuito de demonstrar o funcionamento dos *softwares* e as suas respectivas avaliações. As subdivisões em tópicos são:

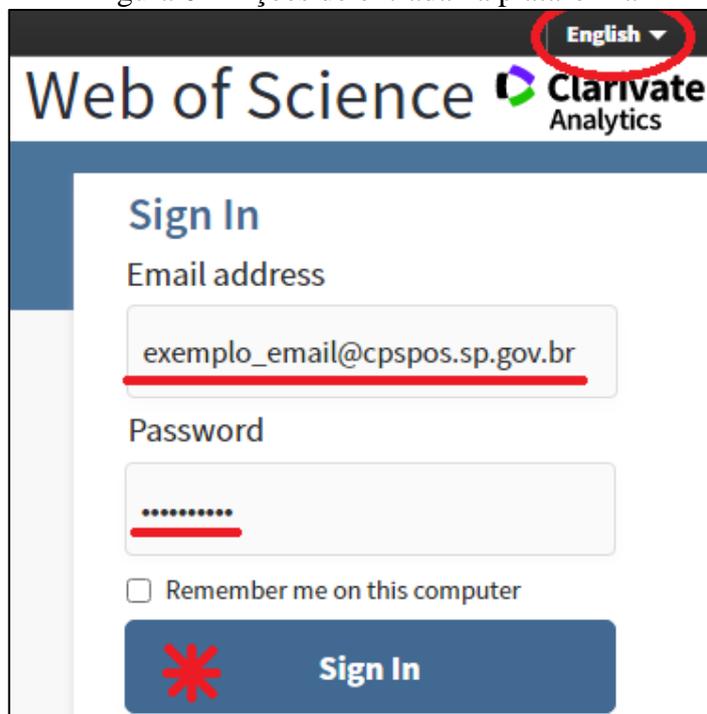
- Funcionamento do robô de busca – *WoS*
- Funcionamento do robô de busca – *Espacenet*
- Funcionamento do robô de busca – *US Patent*
- Avaliação do robô de busca – *WoS*
- Avaliação do robô de busca – *Espacenet*
- Avaliação do robô de busca – *US Patent*

##### 4.1 Funcionamento do robô de busca – *WoS*

Executado o programa, os blocos começam a ser acionados. Seguindo as instruções do código, o *driver* é selecionado e o navegador é aberto.

Com o navegador aberto e o endereço eletrônico da plataforma inserido, as informações de login são inseridas e o idioma é selecionado. A Figura 6 destaca em vermelho as ações executadas para entrar na plataforma.

Figura 6 – Ações de entrada na plataforma



Fonte: Captura de tela durante a execução do programa.

Já dentro da plataforma, o programa seleciona a base de dados em que se quer buscar e o algoritmo de busca, como demonstrado em vermelho na Figura 7.

Figura 7 – Seleção da base de dados e inserção do algoritmo de busca



Fonte: Captura de tela durante a execução do programa.

O programa efetua a busca, abre o primeiro resultado (se houver), capturados dados e caminha pelos resultados encontrados clicando na seta destacada em vermelho.

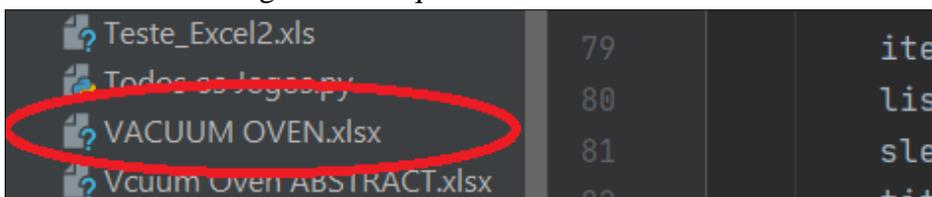
Figura 8 – Navegação capturando dados



Fonte: Captura de tela durante a execução do programa.

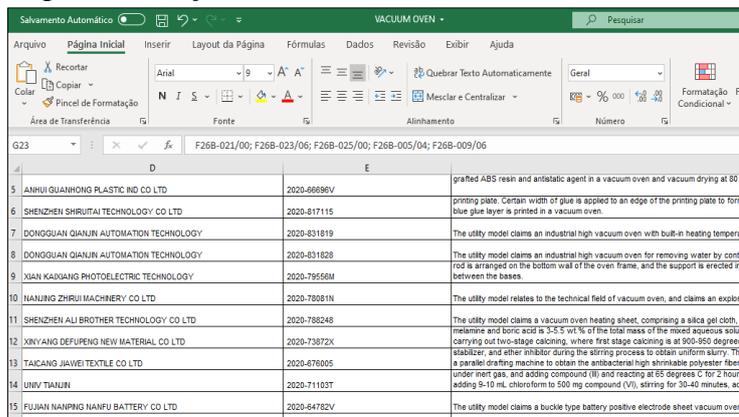
Por fim, após percorrer todas as patentes encontradas, o *software* transforma os dados armazenados em listas em uma matriz e a converte em um documento de planilha eletrônica. A Figura 9 demonstra o arquivo salvo no diretório do ambiente integrador e a Figura 10 demonstra a planilha com os dados aberta.

Figura 9 – Arquivo salvo no diretório



Fonte: Captura de tela no ambiente integrador.

Figura 10- Seção dos dados obtidos e salvos no diretório



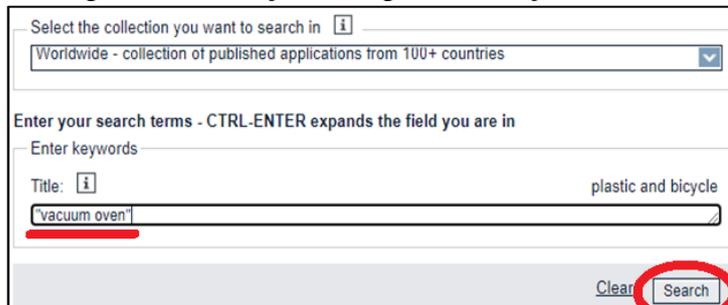
	D	E
5	ANHUI GUANHONG PLASTIC IND CO LTD	2020-86896V
6	SHENZHEN SHRUITAI TECHNOLOGY CO LTD	2020-817115
7	DONGGUAN QIANJIN AUTOMATION TECHNOLOGY	2020-831819
8	DONGGUAN QIANJIN AUTOMATION TECHNOLOGY	2020-831828
9	XIAN KANGANG PHOTOELECTRIC TECHNOLOGY	2020-79558M
10	NINGJING ZHRUI MACHINERY CO LTD	2020-79061N
11	SHENZHEN ALI BROTHER TECHNOLOGY CO LTD	2020-788248
12	XINYANG DEFUPENG NEW MATERIAL CO LTD	2020-73872X
13	TAICANG JIAWEI TEXTILE CO LTD	2020-876005
14	UNIV TIANJIN	2020-71103T
15	FUJIAN NANPING NANFU BATTERY CO LTD	2020-84782V

Fonte: Captura de tela do Excel.

## 4.2 Funcionamento do robô de busca - *Espacenet*

De semelhante forma ao robô de busca do *WoS*, o código seleciona o *driver*, abre o navegador e insere o endereço eletrônico da plataforma. Após essas ações, insere o algoritmo de busca no campo *title* e clica no botão de busca, como mostra a Figura 11 em vermelho.

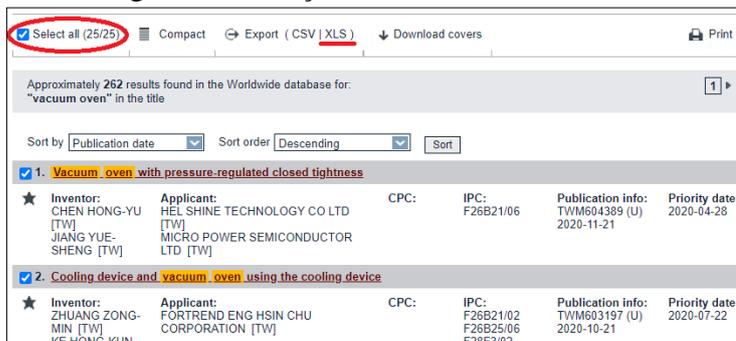
Figura 11- Inserção do algoritmo e ação de busca



Fonte: Captura de tela durante a execução do programa.

O *software* efetua a seleção dos resultados (em looping se necessário em função da quantidade de resultados). Após essa seleção baixa os arquivos em formato de planilha eletrônica, como mostra a Figura 12.

Figura 12- Seleção e *download* dos dados.

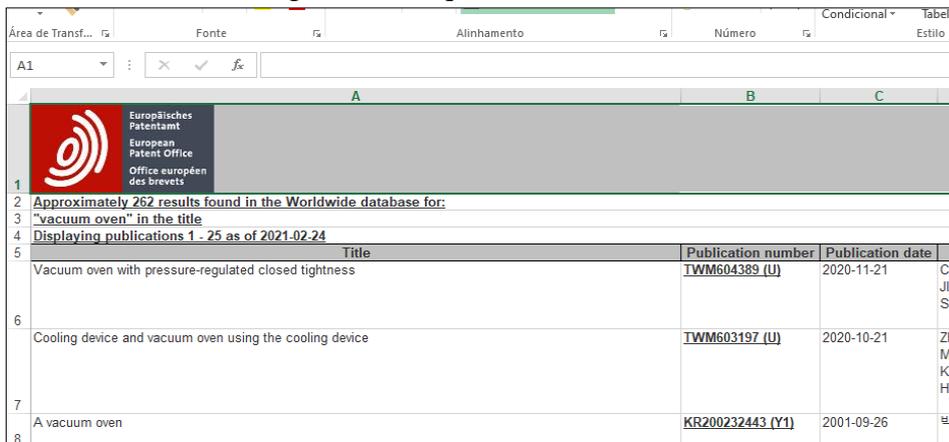


	Inventor:	Applicant:	CPC:	IPC:	Publication info:	Priority date:
1.	Vacuum oven with pressure-regulated closed tightness	CHEN HONG-YU [TW] JIANG YUE-SHENG [TW]	F26B21/06	F26B21/06	TWM604389 (U) 2020-11-21	2020-04-28
2.	Cooling device and vacuum oven using the cooling device	ZHUANG ZONG-MIN [TW] KE HONG-KUN	F26B21/02 F26B25/06 F28E3/02	F26B21/02 F26B25/06 F28E3/02	TWM603197 (U) 2020-10-21	2020-07-22

Fonte: Captura de tela durante a execução do programa.

Após o *download*, o *software* abre o arquivo baixado.

Figura 13- Arquivo baixado aberto



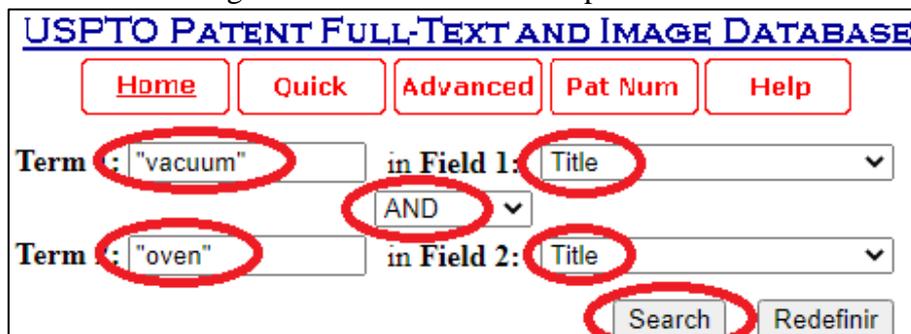
	Title	Publication number	Publication date
5	Vacuum oven with pressure-regulated closed tightness	TWM604389 (U)	2020-11-21
6	Cooling device and vacuum oven using the cooling device	TWM603197 (U)	2020-10-21
7	A vacuum oven	KR200232443 (Y1)	2001-09-26

Fonte: Captura de tela do Excel.

### 4.3 Funcionamento do robô de busca – US Patent

Ao iniciar o programa, o *driver* é selecionado, o navegador aberto e o endereço eletrônico inserido. Após essas ações o algoritmo de busca é inserido, o conectivo de busca selecionado, o filtro de títulos é clicado e o botão de procura também, como mostra a Figura 14 em seus detalhes vermelhos.

Figura 14- Início da busca na plataforma.



**USPTO PATENT FULL-TEXT AND IMAGE DATABASE**  
 Home Quick Advanced Pat Num Help  
 Term : "vacuum" in Field 1: Title  
 AND  
 Term : "oven" in Field 2: Title  
 Search Redefinir

Fonte: Captura de tela durante a execução do programa.

Com a busca efetuada, o programa caminha pelos resultados encontrados capturando os dados das patentes.

Figura 15- Resultados de busca.

PAT. NO.	Title
10.923.720	<a href="#">Graphene-enabled selenium cathode active material for an alkali metal-selenium secondary battery</a>
10.923.716	<a href="#">Graphene-carbon nanotube hybrid electrode material</a>
10.923.567	<a href="#">Graphene FET with graphitic interface layer at contacts</a>
10.923.283	<a href="#">Multilayer ceramic electronic component having an external electrode which includes a graphene platelet</a>
10.920.974	<a href="#">Intelligently-connected vehicle LED headlight using graphene</a>
10.920.045	<a href="#">Elastomeric composition comprising graphene and tire components comprising said composition</a>

Fonte: Captura de tela durante a execução do programa.

For fim, todos os dados até então armazenados em listas são transformados em uma matriz, sendo após salva em formato de planilha eletrônica no diretório do sistema operacional.

#### 4.4 Avaliação do robô de busca – WoS

A primeira etapa de avaliação do *software* que efetua a busca na plataforma do *Web of Science* foi a das qualidades externa e interna com os seus tópicos. O Quadro 4 demonstra a avaliação de qualidade externa e interna do *software*.

Quadro 4 - Avaliação de Qualidade Externa e Interna - WoS

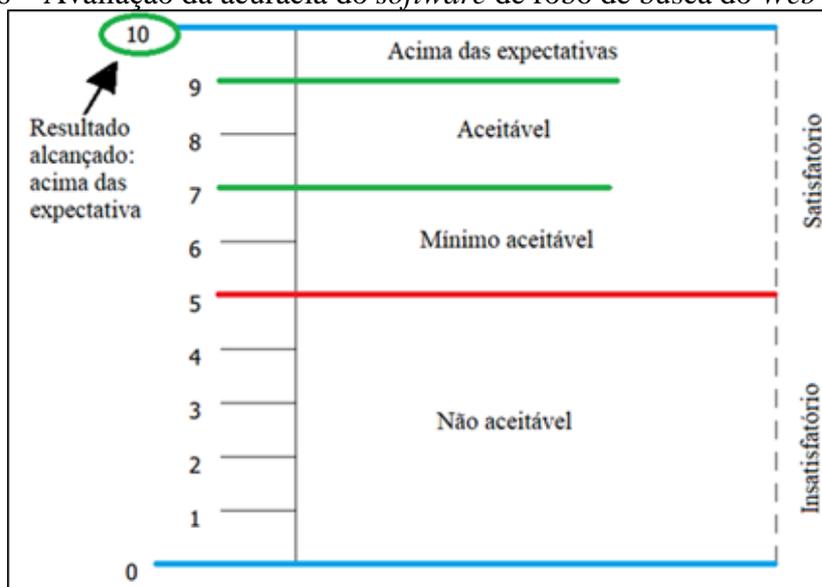
QUALIDADE EXTERNA E INTERNA		
<b>Funcionabilidade</b>	<b>Interoperabilidade</b>	Possui comunicação com outros sistemas e máquinas.
	<b>Acurácia</b>	Demonstra funcionamento no entorno dos pré-requisitos.
	<b>Segurança</b>	Só utilizador manuseia, sem perigo ao equipamento ou à segurança.
	<b>Adequação</b>	Demonstra capacidade de se associar aos objetivos do usuário.
<b>Confiabilidade</b>	Não apresentou falhas durante os testes da acurácia.	
<b>Usabilidade</b>	Fácil compreensão, fácil aprendizado e fácil operação.	
<b>Eficiência</b>	Efetuiu toda a captação dos dados requeridos.	
<b>Manutenibilidade</b>	Pode ser alterado facilmente.	
<b>Portabilidade</b>	Pode usado em outros sistemas operacionais e em outros equipamentos.	

Fonte: O autor.

Quanto à acurácia no funcionamento do *software*, ele se mostra preciso, extraíndo os dados requeridos das páginas da *web*. Para a avaliação da acurácia, foi utilizada a escala de avaliação da Figura 53, medindo a quantidade de vezes que o *software* conseguiu efetuar a retirada de dados e armazená-los até o fim. Os algoritmos utilizados para esses testes foram testados individualmente duas vezes. Foram eles: "*graphene*" AND "*vacuum oven*", "*vacuum oven*", ("*thin films*" OR "*carbon nanotubes*") AND "*vacuum oven*", "*obtaining graphene*", "*graphene production*".

Nas dez operações de busca, o *software* conseguiu captar os dados solicitados e armazená-los como proposto. A acurácia do produto foi medida quantitativamente.

Figura 16 – Avaliação da acurácia do *software* de robô de busca do *Web of Science*



Fonte: os autores.

Na segunda etapa de avaliação foi avaliada a qualidade em uso. O Quadro 5 expressa a avaliação de qualidade em uso do *software* focado nas buscas na plataforma virtual WoS.

Quadro 5 - Avaliação de Qualidade em uso - WoS

QUALIDADE EM USO	
<b>Eficácia</b>	Permite que os usuários atinjam as metas no contexto de uso de busca dentro da plataforma.
<b>Produtividade</b>	O único custo envolvido é o custo do desenvolvedor.
<b>Segurança</b>	Não oferece danos às pessoas, aos negócios ou aos <i>softwares</i> instalados na máquina.
<b>Satisfação</b>	Atendeu às demandas efetuando os trabalhos.

Fonte: os autores.

#### 4.5 Avaliação do robô de busca – *Espacenet*

A primeira etapa da avaliação do *software* que efetua a busca na plataforma do *Espacenet* foi a da qualidade externa e interna com os seus tópicos. O mesmo resultado do Quadro 4 foi o resultado encontrado na avaliação da qualidade externa e interna do *software* robô de busca da plataforma *Espacenet*.

Quanto a acurácia no funcionamento, o mesmo resultado encontrado na medição de acurácia da Figura 16 ocorreu. Os algoritmos utilizados para esses testes foram testados individualmente duas vezes cada. Nas dez operações de busca, o *software* conseguiu captar (minerar) os dados solicitados na busca e armazená-los como proposto, tal como no *software* buscador do WoS. Foram: "vacuum oven", "graphene" AND "vacuum", "obtaining graphene", "graphene production", "thin film treatment".

Na segunda etapa de avaliação do *software*, foi avaliada a qualidade em uso. O mesmo resultado do Quadro 5 foi encontrado durante a avaliação da qualidade em uso do *software* robô de busca que efetua buscas na plataforma *Espacenet*.

#### 4.6 Avaliação do robô de busca – *US Patent*

A primeira etapa da avaliação do *software* que efetua a busca na plataforma do *US Patent* foi a avaliação da qualidade externa e interna com seus subtópicos. O mesmo resultado do Quadro 4 foi o resultado encontrado na avaliação da qualidade externa e interna do *software* robô de busca da plataforma *US Patent*.

Quanto a acurácia no funcionamento, o mesmo resultado encontrado na medição da acurácia da Figura 16 ocorreu. Os algoritmos utilizados para esses testes foram testados individualmente duas vezes. Nas dez operações de busca, o *software* conseguiu captar os dados solicitados e armazená-los como proposto, tal como nos *softwares* buscadores do WoS e da *Espacenet*. Foram usados os mesmos algoritmos utilizados nas buscas da plataforma *Espacenet*.

Na segunda etapa de avaliação do *software*, foi avaliada a qualidade em uso. O mesmo resultado do Quadro 5 foi encontrado durante a avaliação da qualidade em uso do *software* robô de busca que efetua buscas na plataforma *US Patent*.

### 5 Conclusões/Considerações finais

A presente pesquisa atendeu aos objetivos prévios. Foi efetuada a construção de um *software* raspador de dados para a plataforma do *Web of Science*, bem como para as plataformas *Espacenet* e *US Patent*.

O computador demonstrou travamentos em uma busca que continha 60.906 resultados na plataforma da WoS. Este problema foi contornado por meio de buscas parciais, com armazenando fracionado de dados, a saber: 2.000 resultados por busca. A sugestão para buscas futuras é fazer com que as buscas ocorram em uma máquina com especificações mais

adequadas e robustas, com uma velocidade de *internet* e memória de transporte de dados maiores. Seguem as especificações da máquina utilizada na pesquisa:

- Memória de transporte de dados: 4.00 GB;
- Tipo de sistema: Sistema Operacional de 64 bits, processador com base em x64;
- Armazenamento: 120.00 GB;
- Sistema operacional: Windows 10;
- Processador: Intel(R) Core(TM) i7-2670QM CPU @ 2.20GHz 2.20 GHz;
- Rede de internet: 29 Mbps;

As ferramentas construídas possibilitam uma aplicabilidade prática em seu uso, ao mesmo tempo que colabora com o desenvolvimento científico e técnico pela obtenção de informações de difícil obtenção. Utilizando essas ferramentas, é possível fazer diagnósticos de trajetórias tecnológicas dentro das plataformas. Assim, é de importante valia a construção de outros robôs de busca que visem obter dados industriais em outras plataformas de patentes espalhadas pelo mundo.

## Referências

BAYAZIT, N. **Investigating Design: A Review of Forty Years of Design Research, Massachusetts Institute of Technology**. Design Issues, v. 20, n. 1, p. 16-29, 2004. <http://dx.doi.org/10.1162/074793604772933739>.

FALAGAS, M. E., Pitsouni, E. I., MALIETZIS, G. A. e PAPPAS, G. **Scopus, Web of Science, and Google Scholar: strengths and weaknesses**. FASEB J. 22, 338–342 (2008).

FIOCRUZ, BASES DE DADOS PÚBLICOS E BASE DE DADOS PRIVADOS DE PATENES. 2020. Disponível em: < [http://www.castelo.fiocruz.br/vpplr/banco\\_patentes.php](http://www.castelo.fiocruz.br/vpplr/banco_patentes.php)>. Acesso em: 06 de maio de 2020.

HEVNER, A. R.; CHATTERJEE, S. 2010. **Design science in information systems**. Springer.

ISO/IEC 9126. ISO/IEC 9126. Engenharia de software - Qualidade de produto. [S.1.]: **ISO/IEC**, 2003.

ISO/IEC 14598. ISO/IEC 14598: Tecnologia de informação – Avaliação de produto de software. [S.1.]: **ISO/IEC**, 2001.

LABAKI, J. **Introdução a Python – Módulo A**. Universidade Estadual Paulista, 2003.

MENEZES, N. N. C. (2014). **Introdução à Programação com Python**. Editora Novatec, 2ª edição.

MONARD M. C. e BARANAUSKAS J. A. **Conceitos de aprendizado de máquina**. Rezende, Sistemas Inteligentes - Editora Manole, 2003.

RODRIGUES F. C. R (2015). **Manual de Buscas de Patentes**. [Slides do PowerPoint].

SCRAPY, SCRAPY 2.5 DOCUMENTATION. 2021. Disponível em: < <http://https://WWW.docs.scrapy.org/en/latest/>>. Acesso em: 29 de julho maio de 2021.