

Desenvolvimento de busca automatizada de matriz curricular digital, formatadas em PDF e/ou HTML

VICTOR SHIN-ITI OMAE NAKAMURA ¹; BRUNO ANTONY SHIMURA²; GUSTAVO DI CHIACCHIO FAULIN³

¹ Discente em Big Data no Agronegócio na FATEC “Shunji Nishimura”, Pompeia-SP, victorshinitinakamura@gmail.com.

² Discente em Big Data no Agronegócio na FATEC “Shunji Nishimura”, Pompeia-SP.

³ Docente do curso Big Data no Agronegócio, FATEC “Shunji Nishimura”, Pompeia-SP.

RESUMO: O Brasil sendo uma das potências alimentícias do mundo vem avançando na utilização de métodos da Agricultura 4.0, melhorando as técnicas da ciência e tecnologia para otimizar o uso do solo e aumentar a produção, porém surgem desafios para formação de pessoas qualificadas para atuar no ambiente do agronegócio. Com isso foi proposto um estudo para encontrar as universidades que dispõem de temas voltados a agricultura digital e de precisão em seus cursos. Para atingir este objetivo foi desenvolvido dois *crawlers*, um que automatiza a busca por sites de instituições de ensino, e outro por links e arquivos relacionados a matriz curricular dos sites retornados pelo primeiro. Para a construção dos *scripts* dos *crawlers* foi utilizado a biblioteca *Selenium* e o *framework Scrapy*. Os resultados obtidos demonstram que o primeiro *crawler* se mostrou eficiente em seu objetivo, enquanto o segundo se mostrou ineficiente na busca de um a dois sites por vez.

Palavras-chave: Agricultura 4.0. Crawler. Spider.

INTRODUÇÃO

A tendência dos próximos anos é o aumento da demanda por alimentos, ocasionado pelo aumento da população mundial. O Brasil segue entre os maiores produtores e exportadores de *commodities* agrícolas, chegando a ser o segundo maior exportador e produtor de soja, o maior exportador de carnes bovinas e de aves e o maior produtor e exportador de açúcar (ARAGÃO, CONTINI, 2020). Mesmo sendo uma potência alimentícia do mundo e com uma vasta extensão territorial, a capacidade de quanto uma produção consegue se expandir tem um limite. A alternativa é utilizar o melhor das técnicas, da ciência e da tecnologia a fim de otimizar o uso da terra. A agricultura 4.0 emprega métodos e técnicas tanto da informática como rede de sensores, comunicação máquina para máquina, dispositivos móveis, computação em nuvem, internet das coisas, e de Big Data, como soluções analíticas para processar grandes volumes de dados e criar sistemas de suporte de

tomada de decisão, quanto da agricultura e pecuária de precisão, automação e robótica agrícola (MASSRUHÁ, LEITE, 2017, p. 29).

O avanço tecnológico no campo permitiu à agricultura o uso de diversos equipamentos como sensores, máquinas e sistemas integrados que possibilitam aumento do rendimento do solo e maiores colheitas. Apesar da tecnologia ser uma grande aliada, é necessário um profissional capacitado para operá-la, mesmo a empregabilidade sendo alta no estado do Mato Grosso, o setor ainda carecia de mão de obra qualificada (SENAR-MT, 2021). Os profissionais cuja demanda é maior e mais urgente são: técnico em agricultura digital, técnico em agronegócio digital e engenheiro agrônomo digital (FANTIN, 2021).

Segundo um plano de ação realizado pela Câmara do Agro 4.0 (2021), um dos desafios para dispor de recursos humanos qualificados para atuarem no ambiente do Agronegócio, é definir estratégias para a incorporação dos temas da agricultura digital e de precisão nos cursos de graduação e pós-graduação no país. Dito isso, são necessários estudos e formas de automação para identificação de universidades que já possuem, disciplinas relacionadas à agricultura digital e de precisão, em seus cursos.

Uma pesquisa com objetivo de quantificar o número de cursos que oferecem a disciplina de agricultura de precisão, através de contato telefônico, e e-mail realizada pela Piaia (2018), em um total de 69 universidades entrevistadas, 35 cursos oferecem a disciplina, e entre essas, algumas são ofertadas como optativas o que reforça a potencialidade que existe para ampliação da formação agrônômica sobre essa temática nas universidades.

Hoje em dia, as instituições de ensino possuem sites hospedados o que facilitam a busca de informações dentro do site, isso possibilita com que um *crawler/spider* rastreie sites em busca de informações específicas de forma automática, sem necessidade métodos manuais de obtenção de informação demoradas e ineficientes, livres de bloqueio e censura, visto que uma vez estando público não há como barrar.

Os *web crawlers* ou rastreadores são usados desde os anos 90, são essenciais para indexação e classificação das páginas. De acordo com o Google, “O processo de rastreamento começa com uma lista de endereços da web de rastreamentos anteriores e mapas de sites fornecidos pelos proprietários de sites”, no nosso caso, o *crawler* começa consumindo um arquivo CSV, também coletado por outro *crawler*, que contém a lista dos sites a serem pesquisados. Ainda de acordo com Google os *crawlers* funcionam como se fossem pessoas navegando pela página web, entrando de link em link e descobrindo novas páginas que contém referências de mais páginas.

O *crawler* desenvolvido buscará os links para compor um fila de requisições, quando não encontrar o solicitado na página atual, o *crawler* passa para a próxima página na fila, funcionando da mesma forma que o algoritmo de busca em profundidade. Para obter os primeiros links um outro *crawler* será desenvolvido simultaneamente única e exclusivamente para esse propósito.

O *Selenium* (SELENIUM, 2021) é uma ferramenta frequentemente utilizada para testes de aceitação, e raspagem de dados, e torna possível executar ações pré-programadas em páginas web de forma automatizada.

O *Scrapy* (ZYTE, 2008), é um *framework* de rastreamento e raspagem de sites web, usado para rastrear *websites* e extrair a estrutura de dados presente nelas, seja pequenos textos ou estruturas inteiras. Possui diversas funcionalidades úteis na criação do *crawler/spider*, como por exemplo um modelo de *Spider*, *dupefilter* (opção que possibilita o *crawler* não rastreie a mesma página duas vezes), *parser* (transformar os dados), entre outros.

O *Pandas* (MCKINNEY, 2008) usada principalmente em análise de dados e manipulação, porém apenas foi usada para facilitar o uso de arquivos CSV.

O objeto foi desenvolver um *crawler* capaz de buscar páginas relacionadas com a matriz curricular dos cursos de agro, sejam elas páginas HTML ou links de projetos pedagógicos formatadas em PDF.

MATERIAL E MÉTODOS

A linguagem de programação usada foi o *Python* 3.8.8 junto com o *software* de gerenciamento de pacotes *python* Conda 4.10.3, as bibliotecas e *frameworks* instalados foram *Selenium* 4.0.0, *Scrapy* 2.4.1 e *Pandas* 1.2.4. O *hardware* usado no desenvolvimento foi um Intel Core i7-2670QM CPU 2.20GHz × 4, 8 GB de memória RAM e um NVIDIA Corporation GF108M para o desenvolvimento e teste do *crawler*, já para a aplicação com *Selenium* foi utilizado um hardware Intel Core i5-7200U CPU 3.10GHz x 4, 16 GB de memória RAM e um NVIDIA GeForce 920MX, com o sistema operacional Linux Mint 20.1 Cinnamon.

Para encontrar as faculdades com os principais cursos relacionados com agronegócio, utilizamos o site do e-MEC, que retorna todas as instituições em formato de tabela, como visto na Figura 1, porém não retorna o link oficial de cada universidade. Para encontrar esses domínios utilizamos a biblioteca *Selenium* da linguagem *python* realizando assim a busca automatizada no Google de mais de 400 instituições.

Figura 01 - Site oficial e-MEC.

Resultado da Consulta Por : CURSO									
Histórico de Índices Exportar Detalhado Exportar Excel									
Instituição - IES	Sigla	Curso	Grau	Modalidade	Índices	Vagas Anuais	Data Início		
(1) UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	(9) AGRONOMIA	Bacharelado	Presencial	CC: 3(2008) CPC: 4(2019) ENADE: 4(2019) IDD: 3(2019)	86	14/11/1974		
(1) UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	(100778) AGRONOMIA	Bacharelado	Presencial	CC: 3(2011) CPC: 4(2019) ENADE: 4(2019) IDD: 3(2019)	100	07/08/2006		
(1) UNIVERSIDADE FEDERAL DE MATO GROSSO	UFMT	(116662) AGRONOMIA	Bacharelado	Presencial	CC: 3(2012) CPC: 2(2019) ENADE: 3(2019) IDD: 1(2019)	45	02/03/2009		
(2) UNIVERSIDADE DE BRASÍLIA	UNB	(135) AGRONOMIA	Bacharelado	Presencial	CC: 4(2017) CPC: 4(2019) ENADE: 4(2019) IDD: 3(2019)	160	01/03/1962		
(4) UNIVERSIDADE FEDERAL DO AMAZONAS	UFAM	(399) AGRONOMIA	Bacharelado	Presencial	CC: 3(2008) CPC: 4(2019) ENADE: 4(2019) IDD: 4(2019)	46	01/07/1976		

Fonte: <https://emec.mec.gov.br/>

A segunda implementação foi feita simultaneamente com a primeira, na qual um *crawler* é criado usando o *framework Scrapy*. Ele consome o nome do curso, o nome da instituição e o domínio localizado no csv criado pela web *crawler* anterior. Para que o *crawler* pudesse vasculhar por todo o site foi preciso extrair todos os links que encontrasse. Para identificar as páginas que contêm a matriz curricular precisaria possuir dentro de si o nome do curso a ser pesquisado (“Agronomia”, “Engenharia Agrônômica”, etc..) e as palavras-chave (“curricular”, “fluxo de habilitação”, “programa completo”, “itinerário”, “módulo”, etc). Retirando algumas partes da estrutura do HTML como `<header>`, `<footer>`, `<aside>`, `<nav>`, permite que o algoritmo busque apenas no corpo principal da página, pois a div principal nem sempre possui o mesmo nome, às vezes “*main-container*”, outras “*main*”, além de que em certos sites o nome do curso e a palavra-chave estão presentes em todas as páginas que engloba um menu ou navegação, assim o *crawler* salvaria todas elas, o que não é o foco.

A adição de filtros se mostrou necessário, pois há certos sites que possuem links, mas apontam para lugar algum, ou acionam um *javascript*, ou apontam para os elementos da própria página ou são links vazios “#”, um filtro resolve o problema, evitando que o *scrapy* avance para um link problemático, caso contrário o *crawler* entende que não há mais links a ser rastreado e então finaliza. Outro filtro permite que o *crawler* pare de buscar links em urls que contêm um dos conjuntos de palavras-chave (“notícia”, “file”, “revista”, Etc.).

Por fim, caso a página atual atenda todas as condições (potencialmente a página que contém a matriz curricular) ela será salva no formato JSON contendo o nome da instituição, o link dela e o nome do curso.

RESULTADOS E DISCUSSÃO

Na primeira tentativa de otimizar a busca dos sites oficiais das universidades, foi utilizado uma biblioteca Python chamada *googlesearch*, porém após alguns minutos executando o algoritmo, o site de busca Google começou a identificar o *crawler*, pelas quantidades de requisições e a velocidades da pesquisa. O código foi alterado para diminuir o tempo entre as requisições, mas o Google redireciona a pesquisa para uma interface do *CAPTCHA*, que pede para usuários selecionar as imagens relacionadas à palavra exibida na tela, para proteger os sites de robôs tentando acessar áreas restritas.

Depois de algumas pesquisas, a biblioteca *selenium* trouxe melhores resultados, pois o *script* abre o navegador (Browser Firefox), faz requisições mais lentas, simula cada *click* na barra de pesquisa impedindo a identificação do *crawler*. Após pesquisar as 400 universidades, durando por volta de 15 minutos, o algoritmo gera uma tabela no formato CSV com todos os domínios (Figura 2).

Figura 02 - Lista de Instituições com domínio

1	Instituicao	Domínio
2	UNIASSELVI	https://portal.uniasselvi.com.br/
3	UNILASALLE - LUCAS	https://www.unilasalle.edu.br/lucas-do-rio-verde
4	FATEC SENAI MT	https://fatecsenai.com.br/
5	UNIVERSIDADE FEDERAL DE MATO GROSSO	https://cpcs.ufms.br/
6	UNIVERSIDADE DE BRASÍLIA	https://unb.br/
7	UNIVERSIDADE FEDERAL DO AMAZONAS	https://icet.ufam.edu.br/
8	UNIVERSIDADE FEDERAL DE VIÇOSA	https://www.agn.caf.ufv.br/
9	UNIVERSIDADE ESTADUAL DE LONDRINA	http://www.uel.br/
10	PONTIFÍCIA UNIVERSIDADE CATÓLICA DO PARANÁ	https://www.pucpr.br/
11	UNIVERSIDADE DE CAXIAS DO SUL	https://www.uces.br/site
12	UNIVERSIDADE FEDERAL DE UBERLÂNDIA	https://ufu.br/
13	UNIVERSIDADE DE PASSO FUNDO	https://www.upf.br/
14	UNIVERSIDADE ESTADUAL DE SANTA CRUZ	http://www.uesc.br/
15	UNIVERSIDADE VALE DO RIO VERDE	https://www.unincor.br/
16	UNIVERSIDADE JOSÉ DO ROSÁRIO VELLANO	https://www.unifenas.br/
17	UNIVERSIDADE DO ESTADO DA BAHIA	https://portal.uneb.br/
18	FUNDAÇÃO UNIVERSIDADE DO ESTADO DE SANTA CATARIN	https://www.udesc.br/
19	UNIVERSIDADE ESTADUAL DE GOIÁS	http://www.ueg.br/

Fonte: Autoria própria

Com o segundo *crawler*, o teste com os domínios <https://cpcs.ufms.br/>, <http://www.uel.br/>, <https://icet.ufam.edu.br/>, foram bem-sucedidos, o *crawler* varreu por centenas de páginas para achar o conteúdo, ao abrir o CSV, encontra-se alguns dados que não são relevantes, referentes a notícias e afins, mas em todos os testes encontra-se a página desejada (Figura 3).

Figura 03 - JSON retornado dos testes da UFMS e UFAM

```
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO AMAZONAS",
  "curso": "agronomia",
  "url": "https://icet.ufam.edu.br/ensino/graduacao/agronomia/",
  "pdfs": ["http://www.icet.ufam.edu.br/informativos/", "http://edoc.ufam.edu.br/retr"]
},
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO AMAZONAS",
  "curso": "agronomia",
  "url": "https://icet.ufam.edu.br/ensino/graduacao/farmacia/",
  "pdfs": ["https://icet.ufam.edu.br/ensino/graduacao/engenharia-sanitaria/", "https://icet.ufam.edu.br/sistemas-de-informacao/", "http://www.icet.ufam.edu.br/ensino/graduacao/licenciatura-em-ciencias-matematica-e-bio"]
},
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO AMAZONAS",
  "curso": "agronomia",
  "url": "https://icet.ufam.edu.br/ensino/graduacao/quimica-industrial/",
  "pdfs": ["http://icet.ufam.edu.br/ensino/graduacao/engenharia-de-software/", "https://icet.ufam.edu.br/resolucoes/", "http://www.icet.ufam.edu.br/wp-content/uploads/2020/07/certificados-do-programa-bolsa-trabalho-2020.pdf", "https://icet.ufam.edu.br/ufam-oferta-vagas-para-curso-de-musica-na-modalidade-eduacao-a-distancia-2020-1.pdf", "https://icet.ufam.edu.br/page/3/?s=covid", "http://edoc.ufam.edu.br/retr"]
},
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO AMAZONAS",
  "curso": "agronomia",
  "url": "https://icet.ufam.edu.br/comite-de-enfrentamento-ao-coronavirus-emite-nota-tecnica-2020-1.pdf",
  "pdfs": ["https://icet.ufam.edu.br/nova-diretoria-do-icet-toma-posse-para-o-quadrinio-2020-2023.pdf", "https://icet.ufam.edu.br/comunicado-aos-calouros-do-icet-2020-1.pdf", "http://edoc.ufam.edu.br/diretoria-2020/", "http://www.icet.ufam.edu.br/"]
},
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO MATO GROSSO",
  "curso": "agronomia",
  "url": "https://cpcs.ufms.br/docentes/administracao/secretaria-academica/",
  "pdfs": ["https://cpcs.ufms.br/cursos-de-graduacao/engenharia-florestal/", "https://cpcs.ufms.br/cursos-de-graduacao/agronomia-estrutura-curricular-2020.pdf", "https://cpcs.ufms.br/bacharelado-em-administracao-cpcs/", "https://cpcs.ufms.br/cursos-de-graduacao/agronomia/", "https://cpcs.ufms.br/portal-do-calouro/", "https://cpcs.ufms.br/fileadmin/user_upload/2020/07/professores-podem-participar-de-capacitacao-em-empreendedores-2020.pdf", "https://cpcs.ufms.br/2021-tera-24-novos-acordos-com-agentes-de-integracao-parceiros-2021.pdf", "https://cpcs.ufms.br/category/documentos/page/7/", "https://cpcs.ufms.br/estudantes-podem-se-inscrever-nos-grupos-do-programa-de-estudo-2021.pdf", "http://edoc.ufam.edu.br/autor/elian/page/5/", "https://icet.ufam.edu.br/quarta-do-pet-retoma-atividades-e-agora-emite-certi"]
},
{
  "Instituicao": "UNIVERSIDADE FEDERAL DO MATO GROSSO",
  "curso": "agronomia",
  "url": "https://icet.ufam.edu.br/recomposicao-do-nde-da-engenharia-sanitaria/",
  "pdfs": ["https://icet.ufam.edu.br/livro-sobre-pesquisa-no-curso-de-ciencias-farmacuticas-2021.pdf", "http://edoc.ufam.edu.br/revista/2021/04/", "http://edoc.ufam.edu.br/autor/anderson/page/85/", "http://edoc.ufam.edu.br/recomposicao-do-colegiado-do-curso-de-engenharia-sanitaria-2021.pdf", "http://edoc.ufam.edu.br/autor/anderson/page/15/", "http://edoc.ufam.edu.br/autor/elian/page/7/", "http://edoc.ufam.edu.br/next-divulga-edital-2021/", "https://edoc.ufam.edu.br/"]
}
```

Fonte: Autoria própria

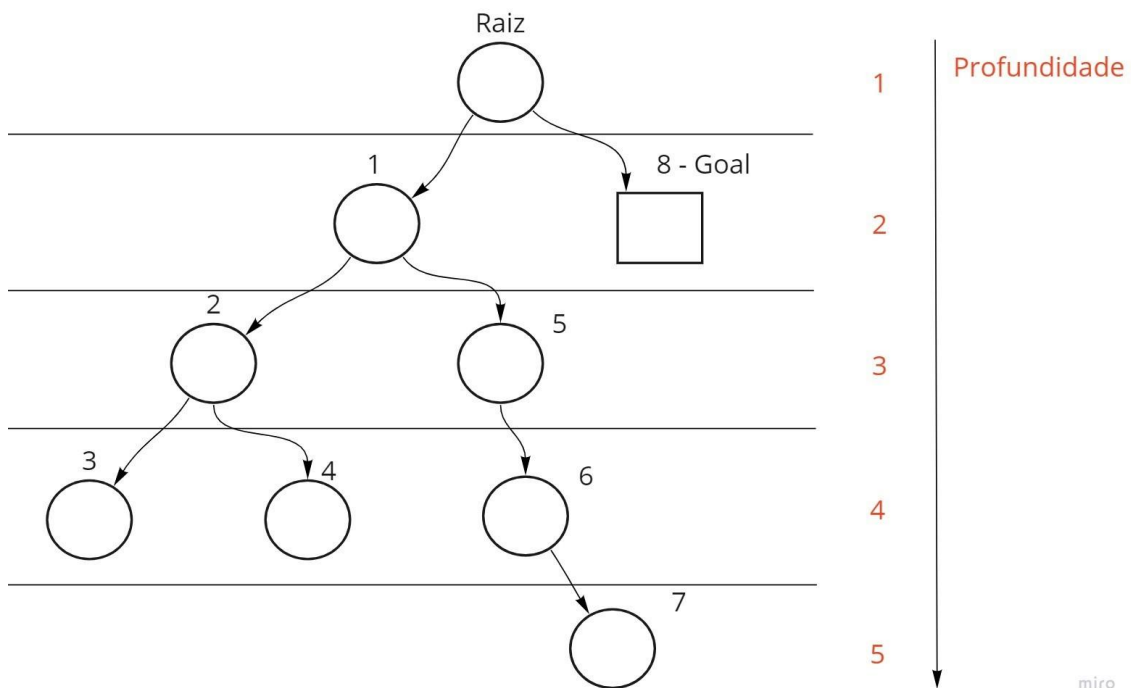
Em alguns sites como do subdomínio de agronomia da Universidade Federal de Viçosa, a página inicial já possuía o nome do curso (Agronomia) e a palavra-chave (Matriz Curricular e Projeto Pedagógico), mesmo com o filtro anti *header*, *footer* e *nav*, pois o conteúdo do menu estava na *tag* `div`, ou seja, o algoritmo salvou todos as páginas do site.

O experimento se mostrou bastante ineficiente caso seja adicionado na fila de urls mais de dois sites diferentes. No primeiro teste, o programa foi rodado com apenas um site em sua fila, enquanto no segundo, foi lido o CSV contendo os domínios do site para ser inserido à fila (cerca de 400 sites). O primeiro teste percorreu 690 páginas em 9 minutos do site da Universidade Federal do Mato Grosso, retornando um arquivo JSON e finalizando o *crawler*, no segundo teste, rodou por 17 horas seguidas, mas apenas percorreu 2519 páginas sem retornar conteúdo algum.

Mais alguns testes foram realizados, rodando com dois *crawlers* ao mesmo tempo, o primeiro rastreando o site da Universidade Federal do Amazonas e o segundo rastreando a Universidade Federal do Mato Grosso, no total foram 8.641 páginas em 1 hora e 30 minutos.

A quantidade de páginas rastreadas por minuto não é fixa, depende muito da topologia do site, a Figura 04 mostra o problema do algoritmo de busca em profundidade, nela cada nó representa um link, o nó quadrado é o link da matriz curricular, o algoritmo percorre o primeiro nó inteiro desnecessariamente para chegar no nó objetivo que estava em uma profundidade menor.

Figura 04 - Topologia Exemplo



Fonte: Autoria Própria

Na etapa de planejamento, o programa foi pensado em usufruir do motor de busca do Google, por ser mais eficiente e demandar menos tempo de execução (o que foi descoberto mais tarde). Também se pensou na construção de um terceiro *crawler*, mas focado em identificar, nas matrizes curriculares, se possuem matérias relacionadas à agricultura 4.0, agricultura de precisão, Big Data e afins. O abandono destes foi por conta do prazo e por fins educativos, a meta foi acima de tudo pôr em prática noções de *scraping* e *crawling*, obter habilidades do uso das ferramentas e experienciar as dificuldades na elaboração de um *crawler*.

CONSIDERAÇÕES FINAIS

O *framework Scrapy* apesar de ser uma boa ferramenta de *web scraping* seu desempenho é insatisfatório na busca simultânea de múltiplos sites, é possível uma melhora no desempenho trabalhando em conjunto com ferramentas de processamento paralelo como o *Hadoop*.

O programa por funcionar a base de listagem, checagem e acesso de links permite seu uso para outros fins que não seja a busca cega de conteúdo, parte de seu código pode ser modificado para atender à outras demandas, como para *Scanner* de site ou para *Backup* de sites.

REFERÊNCIAS

ARAGÃO, Adalberto; CONTINI, Elisio. **O Agro no Brasil e no Mundo: Uma Síntese do Período de 2000 a 2020**. Embrapa SIRE. Brasília, DF:Portal Embrapa (Versão 3.112.0) p01, 2020. Disponível em: <https://www.embrapa.br/documents/10180/62618376/O+AGRO+NO+BRASIL+E+NO+MUNDO.pdf/41e20155-5cd9-f4ad-7119-945e147396cb>. Acesso em: 28 out. 2021.

CÂMARA AGRO 4.0, 2021, Brasília. **PLANO DE AÇÃO DA CÂMARA DO AGRO 4.0**. Disponível em: https://www.gov.br/mcti/pt-br/acompanhe-o-mcti/transformacaodigital/arquivoscamaraagro/ca_plano-de-acao-2021-2024_26-04-2021.pdf. Acesso em: 27 out. 2021.

GOOGLE. **How Search organizes information**. Google Search. Disponível em: <https://www.google.com/search/howsearchworks/crawling-indexing/#:~:text=Finding%20information%20by%20crawling&text=We%20use%20software%20known%20as,webpages%20back%20to%20Google's%20servers>. Acesso em: 28 out. 2021

GOOGLESEARCH. **A Python library for scraping the Google search engine**. Disponível em: <https://pypi.org/project/googlesearch-python/>. Acesso em: 29 out. 2021.

MASSRUHA, Silvia Maria Fonseca Silveira; LEITE, Maria Angélica de Andrade. AGRO 4.0 - rumo à agricultura digital. **JC na escola ciência, tecnologia e sociedade: mobilizar o conhecimento para alimentar o brasil**. São Paulo, 2ª edição, p. 29, 2017. Disponível em: <https://www.agbbauru.org.br/publicacoes/Mobilizar2ed/pdf/Mobilizar2edcompleta.pdf>. Acesso em: 2 nov. 2021

MCKINNEY, WES, et al. **Pandas: a Foundational Python Library for Data Analysis and Statistics**, p. 1, 2008. Disponível em: https://www.dlr.de/sc/Portaldata/15/Resources/dokumente/pyhpc2011/submissions/pyhpc2011_submission_9.pdf. Acesso em: 27 out. 2021.

PIAIA, Manúcia. **O ensino e a pesquisa em agricultura de precisão nos cursos de agronomia nas instituições da região sul do Brasil**. Disponível em: https://repositorio.ufsm.br/bitstream/handle/1/14367/DIS_PPGAGRONOMIA_2018_PIAIA_MANUCIA.pdf?sequence=1&isAllowed=y. Acesso em: 28 out. 2021.

SELENIUM. **Selenium WebDriver**. Disponível em: <https://www.selenium.dev>. Acesso em: 27 out 2021

SENAR-MT. **Profissionais capacitados contribuem para o déficit de mão de obra na mecanização agrícola**. 2021. Disponível em: <https://www.cnabrazil.org.br/noticias/profissionais-capacitados-contribuem-para-odeficit-de-mao-de-obra-na-mecanizacao-agricola>. Acesso em: 28 out. 2021.

UNIVERSIDADE FEDERAL DE LONDRINA. **UEL - Universidade Estadual de Londrina**. Disponível em: <https://portal.uel.br/>. Acesso em: 30 out. 2021

UNIVERSIDADE FEDERAL DE VIÇOSA. **Agronomia | Curso de Agronomia**. 2013. Disponível em: <http://www.agn.ufv.br/>. Acesso em: 29 out. 2021

UNIVERSIDADE FEDERAL DO AMAZONAS. **ICET – Instituto de Ciências Exatas e Tecnologia**. Disponível em: <https://icet.ufam.edu.br/>. Acesso em: 30 out. 2021

UNIVERSIDADE FEDERAL DO MATO GROSSO. **Página Inicial - Campus de Chapadão do Sul**. 2015. Disponível em: <https://cpcs.ufms.br/>. Acesso em: 30 out. 2021

ZYTE, et al. **Scrapy**. 2008. Disponível em: <https://scrapy.org/>. Acesso em: 27 out. 2021.