

Publication status: This preprint has not been published elsewhere.

# APPLICATION OF PYTHON LANGUAGE IN SEARCH ENGINE OPTIMIZATION: EXPLORING ITS CONTRIBUTION TO DATA ANALYSIS

Felipe Ivo da Silva, Gustavo Camossi, Marilde Terezinha Prado Santos, Cecílio Merlotti Rodas

<https://doi.org/10.1590/SciELOPreprints.14513>

Submitted on: 2025-12-10

Posted on: 2025-12-10 (version 1)

(YYYY-MM-DD)

**Journal of Information Systems and Technology Management – Jistem**  
**ISSN online: 1807-1775 - PREPRINT SCIELO**

**APLICAÇÃO DA LINGUAGEM PYTHON NO SEARCH ENGINE OPTIMIZATION:  
EXPLORANDO SUA CONTRIBUIÇÃO PARA A ANÁLISE DE DADOS.**

**APPLICATION OF PYTHON LANGUAGE IN SEARCH ENGINE OPTIMIZATION:  
EXPLORING ITS CONTRIBUTION TO DATA ANALYSIS.**

**Felipe Ivo da Silva**, Universidade Federal de São Carlos, São Carlos, SP, Brasil -  
<https://orcid.org/0009-0005-1379-4692> - e-mail: [felipe.ivo@estudante.ufscar.br](mailto:felipe.ivo@estudante.ufscar.br)

**Gustavo Camossi**, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, SP, Brasil -  
<https://orcid.org/0000-0002-1553-1053> e-mail: [gustavo.camossi@gmail.com](mailto:gustavo.camossi@gmail.com)

**Marilde Terezinha Prado Santos**, Universidade Federal de São Carlos, São Carlos, SP, Brasil -  
<https://orcid.org/0000-0001-5572-2009> - e-mail: [marilde@ufscar.br](mailto:marilde@ufscar.br)

**Cecílio Merlotti Rodas**, Universidade Estadual Paulista Júlio de Mesquita Filho, Marília, SP, Brasil -  
<https://orcid.org/0000-0002-4856-066X> - e-mail: [cecilio.rodas@gmail.com](mailto:cecilio.rodas@gmail.com)

## RESUMO

Este estudo investiga a contribuição da linguagem Python na análise e otimização de dados para Search Engine Optimization (SEO), com foco na organização e recuperação da informação. Adota-se uma abordagem teórico-exploratória, fundamentada em levantamento bibliográfico em bases reconhecidas, como Scopus e Web of Science, além da análise de ferramentas especializadas em SEO. O estudo identificou que Python, por meio de bibliotecas como Pandas e NumPy, permite a automação de processos essenciais, como extração de dados, análise de palavras-chave e modelagem preditiva de padrões de indexação. Os resultados apontam que a aplicação dessas ferramentas melhora a eficiência das estratégias de Search Engine Optimization, tornando-as mais precisas e baseadas em dados. Além disso, destaca-se a interseção entre Ciência da Informação e Ciência da Computação, evidenciando como a programação pode contribuir para a estruturação semântica e organização dos conteúdos digitais. Conclui-se que Python oferece soluções eficazes para a automação e análise preditiva no Search Engine Optimization, potencializando a visibilidade e recuperação da informação nos motores de busca. Sugere-se a realização de estudos experimentais para validar a aplicação prática dessas técnicas em cenários reais. **Palavras-chave: Python; Search Engine Optimization; Análise de dados; Automação.**

## ABSTRACT

This study investigates the contribution of the Python programming language to data analysis and optimization for Search Engine Optimization (SEO), with an emphasis on information organization and retrieval. A theoretical-exploratory approach was adopted, based on a bibliographic review in recognized databases such as Scopus and Web of Science, in addition to the analysis of specialized SEO tools. The study identified that Python, through libraries such as Pandas and NumPy, enables the automation of essential processes, including data extraction, keyword analysis, and predictive modeling of indexing patterns. The results indicate that applying these tools enhances the efficiency of SEO strategies, making them more precise and data-driven. Furthermore, the intersection between Information Science and Computer Science is highlighted, demonstrating how programming can contribute to the semantic structuring and organization of digital content. It is concluded that Python provides effective solutions for automation and predictive analysis in SEO, increasing visibility and information retrieval in search engines. Future research should focus on experimental studies to validate the practical application of these techniques in real-world scenarios.

**Keywords:** Python; Search Engine Optimization; Data analysis; Automation.

## INTRODUÇÃO

A otimização para mecanismos de busca, conhecida pelo termo Search Engine Optimization (SEO), tornou-se fundamental em um cenário digital onde a informação é abundante, mas frequentemente dispersa e desorganizada. A crescente dependência da internet como fonte primária de informações exige que conteúdos digitais sejam estruturados para facilitar sua descoberta, recuperação e interpretação tanto por usuários quanto pelos algoritmos dos mecanismos de busca. Nesse contexto, o SEO se apresenta como um conjunto de técnicas e estratégias que aprimoram a visibilidade de páginas e conteúdos digitais, promovendo um acesso mais eficiente e alinhado às necessidades dos usuários (Camossi; Rodas, 2023).

A eficácia das estratégias de SEO, contudo, depende diretamente de ferramentas capazes de processar grandes volumes de dados, realizar análises detalhadas e automatizar processos. Para isso,

a linguagem de programação Python tem se destacado como uma das mais utilizadas no contexto da análise de dados e automação de processos digitais (Mckinney, 2017). Python é reconhecida por sua simplicidade, ampla gama de bibliotecas e flexibilidade para integração com diversas tecnologias, tornando-se um recurso essencial para profissionais que atuam na otimização de conteúdos digitais. Bibliotecas como Pandas e NumPy oferecem suporte para manipulação, estruturação e modelagem preditiva de dados, conferindo a essa linguagem um diferencial estratégico no desenvolvimento de soluções voltadas para SEO (Vanderplas, 2016).

A interseção entre Ciência da Informação e Ciência da Computação representa um avanço significativo na abordagem dos desafios relacionados ao SEO. Enquanto a Ciência da Informação fornece os fundamentos teóricos sobre a organização, recuperação e classificação de dados, a Ciência da Computação disponibiliza ferramentas tecnológicas necessárias para operacionalizar essas práticas.

Segundo Lewandowski (2023), os mecanismos de busca estão em constante evolução, aprimorando sua capacidade de compreender as intenções dos usuários e hierarquizar conteúdos de maneira mais eficaz. Esse avanço depende não apenas de algoritmos sofisticados, mas também de dados organizados e semanticamente estruturados, que podem ser processados e recuperados por meio da linguagem de programação Python.

Diante desse panorama, este artigo propõe-se a investigar a seguinte questão de pesquisa: qual é a contribuição da linguagem Python na análise e otimização de dados para SEO, especialmente no contexto da organização e recuperação da informação? A relevância deste estudo reside na necessidade crescente de integrar práticas de SEO a tecnologias que possibilitem automação e eficiência no tratamento de dados. A pesquisa busca demonstrar como a aplicação de Python pode contribuir para a análise e otimização de SEO, evidenciando sua interseção com a Ciência da Informação e a Ciência da Computação.

## REFERENCIAL TEÓRICO

### Search Engine Optimization

Search Engine Optimization (SEO) refere-se ao conjunto de técnicas aplicadas para aprimorar a visibilidade de páginas da web nos mecanismos de busca, aumentando a sua relevância nos resultados orgânicos. O SEO pode ser classificado em SEO On-Page, SEO Off-Page e SEO técnico. O SEO On-Page envolve ajustes no próprio site, como a estrutura das URLs, uso de palavras-chave, metadados e otimização de imagens (Heijmans, 2019).

Já o SEO Off-Page foca em fatores externos, como backlinks e menções em redes sociais (Nurkasanah et al., 2022). O SEO técnico, por sua vez, refere-se à infraestrutura do site, incluindo velocidade de carregamento (Dolai, Shenmare e Gudadhe, 2023) e compatibilidade com dispositivos móveis (Xinghai, 2023).

O SEO On-Page é fundamental para garantir que os mecanismos de busca compreendam o conteúdo de um site. Elementos como a escolha de palavras-chave, a densidade e proximidade dessas palavras e o uso adequado de heading tags impactam diretamente o ranqueamento (Search

Engine Journal, 2017). Além disso, a qualidade do conteúdo tem papel determinante, sendo recomendável que os textos tenham um mínimo de 300 palavras para serem indexados de maneira eficiente pelos mecanismos de busca (Rutherford, 2015). Outro aspecto relevante é a otimização das imagens, que devem incluir descrições textuais e alt text para melhor acessibilidade e indexação (Henry et al., 2019).

O SEO Off-Page abrange fatores que aumentam a credibilidade do site, como backlinks e menções externas. Backlinks de qualidade são interpretados pelos algoritmos como um indicativo de confiabilidade e autoridade do site (Brin; Page, 1998). A quantidade e a qualidade desses links são fatores críticos para um melhor ranqueamento, sendo que domínios com alta autoridade costumam gerar impactos positivos mais expressivos (Moz, 2022). Além disso, estratégias de compartilhamento em redes sociais e interações digitais também contribuem para aumentar a visibilidade do site (Sharma e Verma, 2020).

O SEO técnico abrange aspectos estruturais do site que influenciam sua indexação. A velocidade de carregamento é um fator determinante para o ranqueamento, visto que páginas lentas geram altas taxas de rejeição (Kim, 2022). A adoção de técnicas como compressão de imagens, redução do tempo de resposta do servidor e implementação de Content Delivery Networks (CDN) são essenciais para otimizar o tempo de carregamento (Patel e Gupta, 2020). Além disso, o uso de URLs amigáveis e a correta implementação de redirecionamentos e meta tags podem influenciar diretamente a usabilidade e a experiência do usuário (Edgar, 2023).

Outro fator relevante para o SEO é a experiência do usuário (User Experience – UX), que envolve desde a navegabilidade do site até a responsividade para dispositivos móveis (Xinghai, 2023). O design responsivo se tornou um critério de ranqueamento desde a atualização do algoritmo Mobilegeddon do Google, que favorece sites adaptáveis para diferentes tamanhos de tela (Henry et al., 2019). Além disso, a estrutura dos links internos contribui para a distribuição da autoridade das páginas e melhora a navegação dos usuários, favorecendo a retenção e reduzindo a taxa de rejeição (Chen, 2018).

Diante da crescente concorrência digital, aplicar corretamente estratégias de SEO é fundamental para garantir maior visibilidade e relevância nos mecanismos de busca. A integração entre SEO On-Page, SEO Off-Page e SEO técnico proporciona uma abordagem completa para aprimorar o ranqueamento e a experiência do usuário. A evolução constante dos algoritmos de busca reforça a necessidade de atualização e monitoramento contínuos para manter a eficácia das estratégias adotadas (Reisenbichler et al., 2022).

## **Linguagem de programação Python**

A linguagem de programação Python destaca-se no cenário atual da Ciência de Dados e da otimização para mecanismos de busca (Search Engine Optimization – SEO) devido à sua simplicidade, versatilidade e ampla aceitação acadêmica e profissional. Faria (2022), aponta que o Python apresenta uma curva de aprendizado mais acessível em comparação a outras linguagens, o que se justifica pela sua sintaxe intuitiva e pela estrutura de código que se aproxima da linguagem natural.

Do ponto de vista técnico, Python é classificado como uma linguagem de alto nível, multiplataforma e de propósito geral (Eich, 2023). Essa característica multipropósito permite sua

aplicação tanto em tarefas básicas, como manipulação de dados e automação de processos, quanto em áreas complexas como a Inteligência Artificial e o Data Science (Virtus et Veritas, 2024). Especificamente na área de SEO, Python possibilita o desenvolvimento de projetos voltados para a coleta de dados, análise de padrões de busca e automação de auditorias de páginas web, fatores imprescindíveis para o aprimoramento da visibilidade digital.

Diversos autores destacam ainda que o sucesso de Python decorre da sua robusta biblioteca de módulos, com destaque para Pandas, NumPy, Matplotlib e Scikit-learn, amplamente utilizadas na manipulação, visualização e modelagem de dados (Eich, 2023; Virtus Et Veritas, 2024). Estas bibliotecas viabilizam, por exemplo, a execução de tarefas como análise de palavras-chave, identificação de tendências de pesquisa e predição de comportamentos de usuários, ou seja, aplicações diretamente relacionadas à eficácia do SEO moderno.

Além disso, Feltrin (2022), diz que a adoção de Python é impulsionada pela facilidade de integração com bases de dados e pela capacidade de desenvolvimento de soluções escaláveis, fatores críticos no contexto do SEO, em que a análise contínua e a rápida adaptação a novos padrões de indexação se tornaram requisitos estratégicos.

Por fim, é necessário enfatizar que Python, por ser um projeto de código aberto (open source), mantém uma ampla e ativa comunidade de usuários, o que favorece a constante atualização de ferramentas e a disponibilidade de suporte técnico (Ribeiro, 2022). Esta característica proporciona um ambiente propício à inovação, essencial para acompanhar as dinâmicas e evoluções das práticas de SEO em ambientes digitais cada vez mais competitivos.

Dessa forma, a escolha pela linguagem Python para estudos aplicados em SEO não se configura apenas como uma decisão técnica, mas como uma estratégia científica alinhada à busca por eficiência, inovação e democratização do acesso às tecnologias da informação.

O documento esteve sob consulta pública entre janeiro a abril de 2016, passou por consolidação das contribuições e atualizações de maio a junho de 2016 e teve seu Modelo de Referência apresentado no IV Fórum (outubro de 2016), para ficar em consulta pública permanente posteriormente no Participa.br. Por fim, o documento foi validado em Oficina e o V Fórum foi realizado (2017).

## **Estruturas Fundamentais de Python**

A compreensão das estruturas fundamentais da linguagem Python é essencial para sua aplicação eficiente no desenvolvimento de soluções em análise de dados e otimização de informações digitais. Python, enquanto linguagem de programação de alto nível, é caracterizada por uma tipagem dinâmica e forte, uma sintaxe clara e estruturas de controle de fluxo altamente intuitivas (Eich, 2023).

As variáveis em Python, por exemplo, são criadas de forma dinâmica, permitindo a atribuição de diferentes tipos de dados sem necessidade de declaração explícita de tipo, o que proporciona maior flexibilidade no desenvolvimento de algoritmos (Ribeiro, 2022). Essa característica é particularmente vantajosa em aplicações de SEO, nas quais a manipulação de diferentes formatos de dados, como textos, números e listas, é uma constante.

As estruturas de controle, como if-else, for e while, possibilitam a construção de fluxos de decisão e iteração de maneira simples e eficiente (Eich, 2023). O controle condicional é amplamente utilizado em rotinas de verificação de dados em SEO, por exemplo, para validar a conformidade de páginas com boas práticas de otimização ou para realizar auditorias automatizadas de websites.

No que tange às estruturas de dados, Python oferece tipos nativos como listas, tuplas e dicionários, que são fundamentais para o armazenamento e a manipulação de conjuntos de informações (Feltrin, 2022). As listas permitem a organização sequencial de elementos, as tuplas oferecem conjuntos imutáveis de dados, enquanto os dicionários viabilizam a associação de pares chave-valor, sendo este último amplamente utilizado em aplicações que exigem a estruturação de informações extraídas da web, como metadados SEO.

O paradigma de modularização e reutilização de código também é fortemente incentivado em Python, por meio da definição de funções e da utilização de módulos e pacotes (Virtus Et Veritas, 2024). Essa abordagem não apenas favorece a organização e a manutenção de sistemas de análise de SEO, mas também contribui para a escalabilidade de projetos de ciência de dados e automação de tarefas.

Assim, a estrutura básica de Python, que compreende variáveis, operadores, estruturas de decisão, estruturas de repetição, funções e manipulação de dados, fornece as bases indispensáveis para o desenvolvimento de soluções voltadas à análise de dados e à otimização de conteúdos digitais. A simplicidade associada à robustez dessas estruturas torna Python uma linguagem estratégica para a implementação de processos automatizados e análises em larga escala no ambiente digital.

## **Bibliotecas Essenciais para Análise de Dados e SEO com Python**

A efetividade do uso da linguagem Python em projetos de análise de dados e otimização para mecanismos de busca (Search Engine Optimization – SEO) está diretamente relacionada à ampla variedade de bibliotecas especializadas que a linguagem oferece. Essas bibliotecas proporcionam funcionalidades otimizadas para a manipulação, análise, visualização e modelagem de dados, consolidando Python como uma das linguagens mais apropriadas para aplicações em SEO e Ciência da Informação (Eich, 2023).

Entre as bibliotecas essenciais destaca-se o Pandas, que fornece estruturas de dados robustas, como os DataFrames, permitindo a manipulação eficiente de dados tabulares. Essa biblioteca é indispensável para tarefas de limpeza, organização, filtragem e agregação de dados extraídos de websites, fatores fundamentais na construção de análises SEO orientadas por dados (Virtus Et Veritas, 2024).

O NumPy complementa o Pandas ao oferecer suporte para operações de álgebra linear e manipulação de arrays de alta performance, sendo amplamente utilizado para o tratamento de dados numéricos e vetorização de processos analíticos (Eich, 2023). A capacidade de processar grandes volumes de dados de forma eficiente é essencial na avaliação de métricas SEO, como tempo de carregamento de páginas, densidade de palavras-chave e análise de backlinks.

Para a visualização de dados, as bibliotecas Matplotlib e Seaborn desempenham papéis centrais. A Matplotlib é responsável pela criação de gráficos básicos, enquanto o Seaborn,

construído sobre a primeira, permite a elaboração de visualizações estatísticas mais elaboradas e esteticamente refinadas (Ribeiro, 2022). A capacidade de construir gráficos e dashboards é vital para a interpretação de padrões de comportamento de usuários e para a apresentação de relatórios de desempenho SEO de maneira acessível e compreensível.

No campo da modelagem preditiva, a biblioteca Scikit-learn se destaca como ferramenta central para a aplicação de algoritmos de aprendizado de máquina supervisionado e não supervisionado, como regressão, classificação e agrupamento de dados (Virtus Et Veritas, 2024). O uso de técnicas preditivas possibilita, por exemplo, antecipar tendências de tráfego, estimar o impacto de mudanças em páginas web.

Complementarmente, a biblioteca BeautifulSoup é amplamente utilizada para web scraping, permitindo a extração sistemática de informações de páginas web, processo importante para a coleta automatizada de metadados e para a análise competitiva de estratégias de SEO (Feltrin, 2022).

Portanto, o ecossistema de bibliotecas de Python oferece suporte completo a todo o fluxo de trabalho de projetos de SEO baseados em dados, desde a coleta e preparação dos dados, passando pela análise e modelagem, até a visualização dos resultados. A integração dessas bibliotecas constitui um diferencial estratégico para pesquisadores e profissionais que buscam soluções otimizadas e baseadas em evidências no contexto da informação digital.

## **PROCEDIMENTOS METODOLÓGICOS**

Esta pesquisa adota uma abordagem qualitativa, de natureza teórica e exploratória, com o objetivo de analisar como a linguagem Python tem sido aplicada na análise e otimização de dados para Search Engine Optimization (SEO), em especial no contexto da organização e recuperação da informação em ambientes digitais.

A construção da base bibliográfica seguiu um protocolo de busca predefinido, que estruturou os critérios de pesquisa, seleção e extração de dados. O protocolo orientou a identificação de estudos que abordassem a aplicação da linguagem Python na análise e otimização de dados para Search Engine Optimization (SEO), priorizando publicações em periódicos, indexados nas bases Scopus e Web of Science.

Foram utilizados operadores booleanos para a formulação das estratégias de busca, combinando os termos "Python", "Search Engine Optimization" e "Data Analysis". Não se aplicou recorte temporal, e foram considerados apenas documentos publicados nos idiomas inglês e português. A seleção dos estudos foi inicialmente baseada na leitura de títulos e resumos, seguida pela análise integral dos textos elegíveis.

Complementarmente, empregou-se a técnica de snowballing, tanto para explorar as referências dos artigos selecionados quanto para identificar estudos que os citaram, visando garantir a ampliação e a saturação do corpus bibliográfico, conforme metodologia recomendada por Silva (2017).

Para a extração dos dados, foram definidos campos específicos que incluíram: objetivo do artigo, técnicas e bibliotecas de Python utilizadas, tipos de dados e contextos de aplicação, estratégias de SEO abordadas, desafios ou limitações destacados e principais resultados obtidos.

Essas informações subsidiaram a análise qualitativa, possibilitando a identificação de categorias interpretativas alinhadas ao objetivo da pesquisa.

**Quadro 1.** Protocolo de pesquisa

Elemento	Conteúdo
<b>Pergunta de pesquisa (principal)</b>	Como a linguagem Python tem sido discutida e aplicada na análise e otimização de dados para Search Engine Optimization (SEO)?
<b>Objetivos</b>	Identificar como a linguagem Python tem sido utilizada para análise e otimização de dados em SEO, explorando suas aplicações práticas e teóricas no contexto da organização e recuperação da informação.
<b>Palavras-chave</b>	("Python" AND "Search Engine Optimization OR SEO" AND "Data Analysis")
<b>Bases de dados consultadas</b>	<i>Scopus e Web of Science</i>
<b>Estratégia de busca</b>	Uso de operadores booleanos (AND) para combinar termos relacionados ao tema; aplicação de filtros para: artigos de periódicos, artigos de revisão, e anais de conferências.
<b>Período abrangido</b>	Sem recorte temporal.
<b>Tipos de documentos</b>	Artigos
<b>Idiomas</b>	Inglês e português.
<b>CrITÉrios de Inclusão</b>	(I1) Aborda o uso da linguagem Python na análise e otimização de dados voltados para SEO. (I2) Discute a integração de Python com técnicas de organização e recuperação da informação no contexto de SEO. (I3) Apresenta aplicações práticas ou revisões teóricas sobre o tema.
<b>CrITÉrios de Exclusão</b>	(E1) Não aborda diretamente SEO ou a linguagem Python. (E2) Documentos indisponíveis em texto completo. (E3) Trabalhos não escritos em inglês ou português. (E4) Estudo que apenas menciona Python superficialmente sem relação direta com análise de dados para SEO. (E5) Estudo que trata exclusivamente de desenvolvimento web sem conexão com análise de SEO baseada em dados.
<b>Formulário de extração</b>	1) Objetivo do artigo; 2) Técnicas/ferramentas de Python utilizadas; 3) Tipos de dados analisados; 4) Aplicações práticas realizadas (se houver); 5) Desafios e limitações apontadas; 6) Resultados e benefícios reportados.

Fonte: Elaborado pelos autores (2025).

O Quadro 1 foi construído com o intuito de apresentar de forma clara e sistematizada as principais informações extraídas dos estudos selecionados. Foram organizados campos específicos para o registro do objetivo dos artigos, as técnicas e bibliotecas de Python empregadas, os tipos de dados analisados, o contexto de aplicação, as estratégias de SEO abordadas, os desafios e limitações apontados, bem como os principais resultados relatados.

Para o tratamento dos dados, foi empregada a técnica de análise de conteúdo categorial temática, conforme proposta por Bardin (2011), considerada adequada para a identificação de regularidades, núcleos de sentido e padrões interpretativos em um corpus textual heterogêneo. Essa abordagem permitiu a categorização dos estudos analisados em eixos temáticos, contribuindo para a construção de uma análise crítica e fundamentada dos resultados.

Dessa forma, o Quadro 1 funcionou como instrumento fundamental para a organização do material empírico da pesquisa, assegurando transparência, rigor e reprodutibilidade ao processo de análise de conteúdo.

## RESULTADOS E DISCUSSÕES

A aplicação do protocolo de pesquisa resultou na recuperação inicial de quatro estudos relevantes, sendo três oriundos da base de dados Web of Science e um da Scopus. Durante o

processo de triagem, constatou-se que um dos estudos apresentava duplicidade, estando indexado em ambas as bases sob registros distintos, motivo pelo qual foi desconsiderado, conforme a orientação metodológica de exclusão de duplicatas.

Posteriormente, procedeu-se à seleção dos estudos conforme os critérios de inclusão e exclusão estabelecidos no protocolo de pesquisa. Nesse processo, um dos artigos recuperados foi excluído por não atender aos requisitos temáticos e metodológicos: tratava-se de um artigo retratado, sem abordagem direta da linguagem Python no contexto de análise e otimização de dados para Search Engine Optimization (SEO). Assim, a amostra final para análise foi composta por dois estudos únicos e elegíveis, ambos alinhados ao objetivo da pesquisa.

A caracterização dos estudos selecionados, apresentada nos Quadros 2 e 3, evidencia a diversidade de abordagens na utilização da linguagem Python no contexto da análise e otimização de dados para Search Engine Optimization.

**Quadro 2. Aplicações de Python e análise de dados em estudos sobre SEO**

Estudo	Objetivo	Técnicas/Bibliotecas de Python	Tipos de Dados Analisados
Tavosi et al., 2025	Investigar a relação entre a complexidade visual e o ranqueamento SEO de websites de bibliotecas universitárias.	Python AtheC (ferramenta de análise estética computacional)	Dados estruturais de websites (aspectos estéticos e SEO scores)
Cadavid, Valencia-Arias, 2022	Analisar e categorizar a produção científica em marketing digital para PMEs, com ênfase nas tendências de SEO.	Scripts próprios em <i>Python</i> e pacote Bibliometrix no R	Dados bibliométricos de 294 artigos ( <i>Scopus</i> ) sobre marketing digital e SEO

Fonte: Tabela elaborada pelos autores

O Quadro 2 demonstra que ambos os estudos, embora convergentes no uso de Python, divergem quanto à natureza dos dados analisados e à finalidade de suas investigações. No primeiro estudo, realizado por Tavosi et al. (2025), a linguagem Python foi empregada diretamente na análise da estética visual de websites de bibliotecas universitárias, visando identificar possíveis relações entre a complexidade visual e o ranqueamento em mecanismos de busca. Essa aplicação prática da programação reforça a versatilidade do Python para a extração automatizada de indicadores técnicos relevantes para SEO, alinhando-se às práticas de avaliação de desempenho de websites.

Por outro lado, no estudo de Cadavid e Valencia-Arias (2022), a utilização de Python se deu no contexto de uma análise bibliométrica, em que scripts personalizados foram empregados para minerar e processar dados de publicações científicas sobre marketing digital, com destaque para SEO como uma das áreas emergentes. Aqui, observa-se o potencial da linguagem como ferramenta de suporte para análises de larga escala, permitindo mapear tendências e lacunas no campo da informação digital.

A seguir, apresenta-se o Quadro 3, que se aprofunda a compreensão dessas diferenças, ao destacar as aplicações práticas e os resultados principais de cada estudo. No caso de Tavosi et al. (2025), a aplicação de Python centrou-se na coleta e análise de métricas estruturais de SEO técnico, evidenciando a capacidade dos algoritmos para mensurar atributos estéticos e funcionais de websites. O principal destaque deste estudo, é a ausência de correlação significativa entre complexidade visual e ranqueamento em sites de alta credibilidade, isso indica que fatores externos,

como a reputação institucional, podem mitigar a necessidade de estratégias tradicionais de SEO em determinados contextos.

**Quadro 3.** Aplicações práticas e os resultados principais de cada estudo

Estudo	Aplicações Práticas	Estratégias de SEO Abordadas	Principais Resultados
Tavosi et al., (2025)	Avaliação da complexidade visual e extração automática de métricas SEO	SEO técnico (fatores estruturais e de conteúdo em sites institucionais)	Uso inovador de Python para medir atributos estéticos; reforço de que bibliotecas universitárias atraem usuários independentemente do SEO.
Cadavid; Valencia-Arias (2022)	Mineração de dados bibliográficos e análise de coocorrência de palavras-chave	SEO como parte das tendências emergentes no marketing digital	Identificação de SEO como tema consolidado no marketing digital para PMEs; proposta de agenda futura de pesquisa sobre SEO e big data.

Fonte: Elaborada pelos autores

Em contrapartida, o estudo de Cadavid e Valencia-Arias (2022) ressaltou o SEO como uma prática consolidada no marketing digital para pequenas e médias empresas, identificando-o como um eixo estratégico fundamental no ambiente competitivo contemporâneo. A mineração de dados bibliográficos permitiu não apenas a identificação de tendências, mas também a proposição de diretrizes para futuras investigações, como a necessidade de aprofundar o entendimento sobre a integração entre SEO, big data e estratégias de visibilidade digital.

Esses resultados convergem com a literatura revisada, que aponta o Python como uma linguagem-chave tanto para tarefas operacionais, como a automação de auditorias de SEO, quanto para análises meta-analíticas em grandes volumes de dados Tavosi et al., (2025) e Cadavid; Valencia-Arias (2022). Além disso, reforçam a noção de que o SEO transcende a mera otimização de palavras-chave, envolvendo aspectos estéticos, estruturais e estratégicos da organização da informação na web.

Dessa forma, os dados obtidos, sistematizados nos Quadros 2 e 3, permitem afirmar que a linguagem Python se consolida como uma ferramenta multifuncional no ecossistema digital atual, atendendo tanto a demandas técnicas quanto analíticas no âmbito do SEO e da organização da informação.

A análise dos dois estudos selecionados revelou abordagens distintas, porém complementares, acerca da aplicação da linguagem Python no contexto da análise e otimização de dados voltados para Search Engine Optimization (SEO).

O estudo de Tavosi et al. (2025) apresenta uma aplicação direta e prática da linguagem Python por meio da utilização da ferramenta AtheC, desenvolvida especificamente para a avaliação computacional da estética visual de websites. A pesquisa concentrou-se na análise da relação entre a complexidade visual de websites de bibliotecas universitárias e seus respectivos ranqueamentos em SEO. Apesar de não terem encontrado correlação estatisticamente significativa entre os fatores analisados, os autores demonstraram a capacidade da linguagem Python de operacionalizar métricas complexas de estética digital em estudos de SEO técnico. A abordagem evidencia como fatores relacionados à credibilidade institucional podem relativizar a necessidade de técnicas agressivas de otimização em determinados contextos, como o acadêmico.

Por sua vez, o estudo de Cadavid e Valencia-Arias (2022)<sup>3</sup> adota uma perspectiva mais macro analítica e bibliométrica. Utilizando scripts personalizados em Python e o pacote Bibliometrix em

R, os autores realizaram a mineração e a categorização de uma ampla base de artigos científicos relacionados a marketing digital em pequenas e médias empresas. Embora o SEO não tenha sido o único foco da investigação, ele emergiu como uma temática consolidada nas tendências de marketing digital. A análise evidenciou o crescente interesse acadêmico pela integração de SEO com big data e outras tecnologias digitais, além de propor lacunas e oportunidades para futuras pesquisas. Essa abordagem demonstrou a capacidade do Python para o processamento automatizado de grandes volumes de dados bibliográficos e para a extração de padrões relevantes para a ciência da informação e marketing.

Em termos comparativos, observa-se que, enquanto Tavosi et al. (2025) empregam Python diretamente na coleta e análise de atributos técnicos específicos de SEO (fatores de ranqueamento e estética visual de websites), Cadavid e Valencia-Arias (2022) utilizam Python como suporte metodológico para análises meta-analíticas, evidenciando tendências de pesquisa e mapeando o campo de estudos sobre SEO. Ambos os estudos, entretanto, corroboram a premissa de que a linguagem Python constitui uma ferramenta poderosa tanto para a análise operacional quanto para a análise estratégica em SEO.

Além disso, as duas investigações destacam desafios metodológicos relevantes. Tavosi et al. (2025) identificaram limitações relacionadas à generalização dos resultados devido à especificidade da amostra (websites de instituições de alta credibilidade), enquanto Cadavid e Valencia-Arias (2022) apontaram dificuldades na padronização semântica e na categorização de termos em bases de dados extensas, além da limitação decorrente da utilização de uma única fonte de dados (Scopus).

Esses achados, em conjunto, reforçam a importância da contínua atualização das ferramentas e metodologias baseadas em Python, tanto para a avaliação prática de parâmetros técnicos de SEO quanto para a identificação e análise de tendências emergentes na organização e recuperação da informação digital.

## CONCLUSÕES

Este estudo teve como objetivo investigar a contribuição da linguagem Python para a análise e otimização de dados no âmbito do Search Engine Optimization (SEO), especialmente sob a perspectiva da organização e recuperação da informação. Com base na pesquisa bibliográfica e na análise dos estudos selecionados, constatou-se que Python, por meio de seu ecossistema de bibliotecas, oferece soluções eficazes para a automação de processos, a análise preditiva e a manipulação estruturada de dados em projetos de SEO.

A pesquisa revelou que bibliotecas como Pandas, NumPy, Matplotlib, Scikit-learn e BeautifulSoup desempenham papéis centrais na coleta, no processamento e na análise de informações relevantes para estratégias de SEO. Essas ferramentas, ao possibilitarem a análise sistemática de grandes volumes de dados e a modelagem de tendências, contribuem para a formulação de práticas mais eficientes e fundamentadas, fortalecendo a visibilidade e a encontrabilidade de conteúdos digitais.

Os resultados indicam que a utilização de Python se aproxima dos princípios de organização e recuperação da informação, promovendo a estruturação semântica dos dados e otimizando o processo de indexação pelos mecanismos de busca. A interseção entre Ciência da Computação e Ciência da Informação, nesse contexto, apresenta-se como um campo promissor para o desenvolvimento de novas abordagens analíticas aplicadas ao SEO.

Como limitação desta pesquisa, destaca-se a natureza teórico-exploratória da análise, restrita a estudos bibliográficos. A ausência de experimentação prática impede a verificação empírica da eficácia das técnicas de Python em cenários operacionais de SEO, o que sugere cautela na generalização dos resultados.

Para pesquisas futuras, recomenda-se a realização de estudos experimentais que testem, em ambientes reais, a aplicação de scripts em Python para auditorias SEO, análise de palavras-chave, monitoramento de concorrência e predição de ranqueamentos. Sugere-se ainda a ampliação do escopo para incluir diferentes mecanismos de busca e contextos multilíngues, bem como a análise do impacto da evolução dos algoritmos de busca sobre a eficácia das estratégias automatizadas.

Conclui-se que a linguagem Python se consolida como um recurso estratégico para o aprimoramento do SEO, ao integrar eficiência tecnológica com os princípios da Ciência da Informação e Ciência da Computação, fortalecendo a organização, a acessibilidade e a recuperação de informações na web.

## REFERÊNCIAS

- Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, 30(1–7), 107–117.
- Camossi, G., & Rodas, C. M. (2023). Contribuições das técnicas de Search Engine Optimization para encontrabilidade da informação. *Encontros Bibli*, 28, e93371. Recuperado de <https://www.scielo.br/j/eb/a/TBfVWBGq3Wwckkyy69SHqvn/?lang=pt>
- Chen, H. (2018). An analysis of internal linking structures and their impact on SEO performance. *Journal of Digital Information Management*, 16(4), 245–256.
- Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: SAGE Publications.
- Dolai, P., Shenmare, P., & Gudadhe, S. (2023). A comprehensive review of technical SEO techniques and their impact on search engine rankings. *International Journal of Web Studies*, 12(3), 37–49.
- Edgar, W. (2023). The role of structured metadata in search engine indexing. *Information Retrieval Journal*, 26(2), 112–126.
- Eich, L. A. (2023). *Python data science: Do básico ao intermediário*. Amazon Brasil.
- Faria, E. (2022). *Estrutura de dados e algoritmos: Do pseudocódigo ao JavaScript e Python*. Independente.
- Feltrin, F. (2022). *Python na prática: Aprenda linguagem Python através de um projeto real*. Amazon Brasil.
- Géron, A. (2019). *Hands-on machine learning with Scikit-Learn, Keras and TensorFlow* (2nd ed.). O'Reilly Media.
- Gil, A. C. (2019). *Métodos e técnicas de pesquisa social* (7ª ed.). São Paulo, SP: Atlas.
- Heijmans, M. (2019). The impact of keyword placement on on-page SEO rankings. *Journal of Digital Marketing Research*, 5(2), 134–150.
- Henry, R., Marshall, T., & Gregson, S. (2019). Image optimization techniques for improving SEO and accessibility. *Web Technologies Review*, 7(1), 53–67.
- Hu, Y., Deng, B., & Peng, F. (2016). Autoscaling prediction models for cloud resource provisioning. In *Proceedings of IEEE International Conference on Computer and Communications (ICCC)* (pp. 1364–1369). Chengdu, China. <https://doi.org/10.1109/CompComm.2016.7924927>

- Kim, Y. (2022). The effects of website speed on user behavior and search engine rankings. *Computational Marketing Journal*, 10(1), 22–35.
- Lakatos, E. M., & Marconi, M. A. (2017). *Fundamentos de metodologia científica* (7ª ed.). São Paulo, SP: Atlas.
- Lima, R. R., Fernandes, A. M. R., Bombsar, J. R., Silva, B. A., Crocker, P., & Leithardt, V. R. Q. (2022). An empirical comparison of Portuguese and multilingual BERT models for auto-classification of NCM codes in international trade. *Big Data and Cognitive Computing*, 6(1), 8. <https://doi.org/10.3390/bdcc6010008>
- McKinney, W. (2017). *Python for data analysis: Data wrangling with Pandas, NumPy, and IPython* (2nd ed.). Sebastopol, CA: O'Reilly Media.
- Moghaddasi, H., Mohammadpour, A., Bouraghi, H., Azizi, A., & Mazaherilaghagh, H. (2018). Hospital information systems: The status and approaches in selected countries of the Middle East. *Electronic Physician*, 10(5), 6829–6835. <https://doi.org/10.19082/6829>
- Moz. (2022). Search engine ranking factors: An industry study. *Moz Research Reports*, 4(1), 1–27. Recuperado de <https://moz.com>
- Nurkasanh, I., Wahyuni, R., & Alfarisi, R. (2022). The influence of backlink profiles on SEO performance: A case study of e-commerce websites. *Journal of Web Analytics*, 15(4), 87–101.
- Patel, N., & Gupta, P. (2020). Optimizing website speed for SEO and user engagement. *International Journal of Internet Marketing*, 11(3), 42–58.
- Reisenbichler, M., Keller, L., & Schmidt, T. (2022). The evolution of search engine algorithms: Implications for SEO strategies. *Digital Marketing Insights*, 9(2), 98–112.
- Ribeiro, T. P. (2022). *Python: Guia prático de programação – uma abordagem simples, para verdadeiros iniciantes*. Produtividade Programada.
- Rutherford, J. (2015). Keyword density and its impact on search rankings: A meta-analysis. *Journal of Online Search Strategies*, 8(1), 31–47.
- Search Engine Journal. (2017). On-page SEO best practices: Industry recommendations. *Search Engine Journal Reports*, 2(3), 20–39. Recuperado de <https://searchenginejournal.com>
- Sharma, A., & Verma, R. (2020). Social media engagement and its effect on search rankings. *Journal of Digital Media Research*, 6(2), 59–73.
- Shirvani, M. H., Rahmani, A. M., & Sahaf, A. (2020). A survey study on virtual machine migration and server consolidation techniques in DVFS-enabled cloud datacenter: Taxonomy and challenges. *Journal of King Saud University – Computer and Information Sciences*, 32, 267–286. <https://doi.org/10.1016/j.jksuci.2018.07.001>
- Soualhia, M., Fu, C., & Khomh, F. (2019). Infrastructure fault detection and prediction in edge cloud environments. In *Proceedings of ACM/IEEE Symposium on Edge Computing (SEC)* (pp. 222–235). Nova York, NY. <https://doi.org/10.1145/3318216.3363305>
- Souza, F. D., & Souza, J. B. O. (2021). Sentiment analysis on Brazilian Portuguese user reviews. *arXiv*. Recuperado de <https://arxiv.org/pdf/2112.05459.pdf>
- VanderPlas, J. (2016). *Python data science handbook: Essential tools for working with data*. Sebastopol, CA: O'Reilly Media.
- Virtus et Veritas. (2024). *Python na prática da inteligência artificial: Do zero às aplicações avançadas*. Independente.
- Wilmott, P. (2019). *Machine learning: An applied mathematics introduction*. Panda Ohana Publishing.
- Xinghai, L. (2023). Mobile-first indexing and its impact on SEO rankings. *Web Optimization Review*, 14(1), 77–90.

### **Declaração de contribuição dos autores**

#### **Felipe Ivo da Silva**

Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision Validation, Visualization, Writing – original draft Writing – review & editing

#### **Gustavo Camossi**

Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision Validation, Visualization, Writing – original draft Writing – review & editing

#### **Marilde Terezinha Prado Santos**

Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision Validation, Visualization, Writing – original draft Writing – review & editing

#### **Cecílio Merlotti Rodas**

Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Supervision Validation, Visualization, Writing – original draft Writing – review & editing

### **Declaração de conflito de interesse**

Os autores declaram que não há conflito de interesse.

### **Declaração de disponibilidade de dados da pesquisa**

Todo o conjunto de dados de apoio aos resultados deste estudo foi publicado no próprio artigo.



This preprint was submitted under the following conditions:

- The authors declare that the necessary Terms of Free and Informed Consent of participants or patients in the research were obtained and are described in the manuscript, when applicable.
- The authors declare that the preparation of the manuscript followed the ethical norms of scientific communication.
- The authors declare that they are aware that they are solely responsible for the content of the preprint and that the deposit in SciELO Preprints does not mean any commitment on the part of SciELO, except its preservation and dissemination.
- The authors declare that the data, applications, and other content underlying the manuscript are referenced.
- The deposited manuscript is in PDF format.
- The authors declare that the research that originated the manuscript followed good ethical practices and that the necessary approvals from research ethics committees, when applicable, are described in the manuscript.
- The authors declare that once a manuscript is posted on the SciELO Preprints server, it can only be taken down on request to the SciELO Preprints server Editorial Secretariat, who will post a retraction notice in its place.
- The authors agree that the approved manuscript will be made available under a [Creative Commons CC-BY](#) license.
- The submitting author declares that the contributions of all authors and conflict of interest statement are included explicitly and in specific sections of the manuscript.
- The authors declare that the manuscript was not deposited and/or previously made available on another preprint server or published by a journal.
- If the manuscript is being reviewed or being prepared for publishing but not yet published by a journal, the authors declare that they have received authorization from the journal to make this deposit.
- The submitting author declares that all authors of the manuscript agree with the submission to SciELO Preprints.