
**DESENVOLVIMENTO DE FERRAMENTA SEO UTILIZANDO IA COM USO DE DADOS DOS
BUSCADORES PARA MELHORIA DO RANQUEAMENTO NOS BUSCADORES:
ESTUDO DE CASO DO SITE DHILUF TECH*****DEVELOPMENT OF SEO TOOL USING AI WITH DATA USE OF SEARCHERS TO IMPROVE THE
SEEDING IN THE SEARCHERS: CASE STUDY OF THE SITE DHILUF TECH***

Dhiego Costa Oliveira¹, Felipe de Oliveira C. de Sousa², Lucas Lira B. de Souza³, Erwin Alexander Uhlmann⁴

RESUMO: O artigo em questão apresenta o processo de desenvolvimento de uma ferramenta capaz de aplicar de forma automática o conjunto de técnicas *Search Engine Optimization* (SEO) utilizando Inteligência Artificial (IA). O objetivo do desenvolvimento é ter um *software* com capacidade de produção de textos e semântica das frases, com o intuito de diminuir o impacto financeiro das empresas, melhorar o desenvolvimento de conteúdo do *site*, diminuir o tempo de atualização desse conteúdo e melhorar os textos produzidos. A metodologia empregada no artigo compreendeu pesquisas sobre tecnologias *WEB*, as práticas SEO, aplicação do mesmo sobre uma página *WEB*, o uso e aplicação da Inteligência Artificial e análise dos resultados.

PALAVRAS-CHAVE: SEO. Inteligência Artificial. Ranqueamento. Buscadores

ABSTRACT: *The article in question presents the process of developing a tool capable of automatically applying the set of Search Engine Optimization (SEO) techniques using Artificial Intelligence (AI). The purpose of the development is to have software with the ability to produce texts and sentence semantics, in order to reduce the financial impact of companies, improve the content development of the site, shorten the time to update this content and improve the texts produced. The methodology used in the article included research on WEB technologies, SEO practices, application of the same on a WEB page, the use and application of artificial intelligence and analysis of results.*

KEYWORDS: *SEO. Artificial intelligence. Ranking. Search engines*

1 Graduando em Ciência da Computação - 8º semestre. Universidade UNIVERITAS/UNG. dhiego1122@gmail.com

2 Graduando em Ciência da Computação - 8º semestre. Universidade UNIVERITAS/UNG. felipeoc.sousa@gmail.com

3 Graduando em Ciência da Computação - 8º semestre. Universidade UNIVERITAS/UNG. lucasllbgs@gmail.com

4 Orientador. Analista e Desenvolvedor de Sistemas, Arquiteto e Urbanista, especialista em Administração de Sistemas de Informação, Marketing e Docência, Mestre em Análise em Geociências. Universidade UNIVERITAS/UNG. euhlmann@prof.ung.br

INTRODUÇÃO

Desde que a Era da Internet começou vemos o quanto tornou-se indispensável ao ser humano, é usada para comunicação, expressar opiniões, relacionar-se, comprar e etc.

Segundo pesquisa realizada pelo IBGE em 2016, cerca de 67% da população brasileira está conectada à internet. Ainda segundo dados do serviço de proteção ao crédito (SPC) e da confederação Nacional de Dirigentes Lojistas (CNDL) em 2017 cerca de 89% dos usuários fizeram pelo menos uma compra online em 12 meses. A pesquisa nos mostra que 58% destes consumidores foram motivados por preços mais baixos e 45% pela comodidade de comprar sem sair de casa, dessa forma podemos entender que metade das compras realizadas em 1 ano foram pela internet.

Os investimentos neste setor têm crescido gradativamente e o posicionamento nos buscadores é indispensável aos *e-commerces*. Pesquisa realizada pela própria Google 94% dos usuários não passam para a segunda página, e ainda na mesma pesquisa 32,5% dos cliques são realizados no primeiro *link*, enquanto apenas 2,4% pertencem a décima posição.

Há muitas maneiras de se posicionar nos buscadores, através de *links* de redirecionamento, números de visitantes e otimização que nada mais é que preparar o *site* para ser encontrado com facilidade utilizando o SEO.

SEO é um conjunto de técnicas de otimização para páginas *WEB*, essa otimização busca alcançar uma boa classificação dos *sites* nos buscadores da internet como Google, Bing, Yahoo, DuckDuckGo, Uol e etc.

O mercado oferece diversas empresas e profissionais especializados neste assunto com a finalidade de aplicar essas técnicas em *sites* para melhorar a sua classificação, mas não é tão simples como parece, segundo especialistas pode demorar meses para alcançar um bom posicionamento nos buscadores, pois depende muito de como o *site* foi estruturado e além disso é necessário um acompanhamento diário e atualizações constantes para que seu *site* se mantenha em uma boa posição a fim de ultrapassar o concorrente.

A proposta inicial deste artigo é fornecer uma ferramenta que empregue de forma automática a produção de texto para aplicar no SEO, dessa forma como consequência melhorar o ranqueamento nos buscadores da internet utilizando Inteligência Artificial para ler e armazenar frases e palavras-chaves para produzir novos textos com o intuito de diminuir o impacto financeiro das empresas, melhorar o desenvolvimento do conteúdo no *site*, diminuir o tempo de atualização desse conteúdo, melhorar a semântica dos textos produzidos e trazer resultados mais relevantes para os usuários.

MÉTODOS

Para o desenvolvimento do sistema que atuará na *WEB* será necessário a breve explanação de algumas tecnologias, ferramentas e métodos dos quais serão utilizados na implementação do sistema, tais como o conhecimento sobre o que é um *WEB browser*, busca orgânica e paga, *crawler*, *Search Engine Optimization*, *WEB scraping*, linguagem de programação, python, Programação de Linguagem Natural (PLN), *Natural Language Toolkit* (NLTK), Inteligência Artificial e sistema de gerenciamento de banco de dados.

Para maior entendimento do processo que o *software* executará é necessário o entendimento prévio sobre o que é um *WEB browser*, segundo SILVA (2002) *WEB browser* é uma aplicação que permite o processamento de hipermídia, que funciona do lado do cliente do sistema que se comunica através do protocolo de transferência de hipertexto HTTP com um servidor.

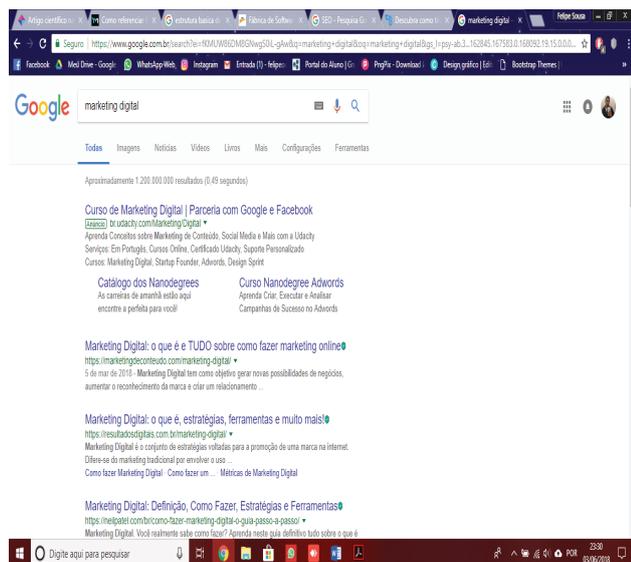
O uso habitual dos *browsers* e para fins de pesquisa de *sites* e um dos objetivos do sistema é a melhora no posicionamento das buscas através do *browser* sendo assim é necessário entender os tipos de buscas.

Existem dois principais tipos de buscas efetuada pelo *browser* sendo elas as buscas orgânicas e pagas. Para DOS SANTOS (2014) a busca orgânica é derivada da análise que o *WEB browser* faz mediante as seguintes considerações, relevância dos *sites* e as palavras-chaves ou *keywords*. Segundo GABRIEL (2010) existem alguns critérios para a busca sendo essa relevância do *site*, *keywords* e o valor do *link* ofertado pelo anunciante esse valor é debitado toda vez que o *link* for clicado gerando custo para o anunciante. Nesse artigo terá enfoque apenas nas buscas orgânicas. Sabendo disso será necessária a análise de como o *browser* retira as informações necessárias para a apresentação do *site* nas buscas.

No exemplo a seguir temos uma pesquisa feito no buscador Google com a palavra marketing digital, no primeiro resultado podemos identificar que se trata de um resultado pago por conta do ícone "anuncio" ao lado esquerdo do *site*, e os demais resultados que trata de um resultado orgânico.

DESENVOLVIMENTO DE FERRAMENTA SEO UTILIZANDO IA COM USO DE DADOS DOS BUSCADORES PARA MELHORIA DO RANQUEAMENTO NOS BUSCADORES: ESTUDO DE CASO DO SITE DHILUF TECH
Diego Costa Oliveira, Felipe de Oliveira C. de Sousa, Lucas Lira B. de Souza, Erwin Alexander Uhlmann

Figura 1 - exemplo de resultados pagos e orgânicos.



O funcionamento de um *WEB browser* é possível a partir da utilização de um *crawler*. Para MENESES (2010) o *WEB crawler* é um *software* que baixa as páginas *WEB* com a finalidade de ser usado por um motor de busca ou cache, no qual, retorna para o navegador a página baixada que salva e cria um índice e analisa o conteúdo das páginas dos *sites*. A partir dessa análise do *crawler* será aplicada métricas de definição a relevância dos *sites* e posicioná-los segundo as métricas de ranqueamento, e para melhorar esse posicionamento do ranque é utilizada um conjunto de técnicas denominadas SEO.

Para GABRIEL (2010) otimização *on-page* SEO é um conjunto de todas técnicas para manipulação do conteúdo que é composta por códigos e estrutura da página *WEB*, que visa a melhoria no posicionamento ranqueado de um *site* nos buscadores através das buscas. O SEO possui algumas características fundamentais para seu funcionamento que constitui de análise do funcionamento nos buscadores e estratégia de produção de conteúdo, código e estrutura da página *WEB* com enfoque no SEO. Essa construção da página e código é dada através do uso da linguagem de marcação HTML.

Segundo os doutores ALVARENGA e SOUZA (2004) O *Hypertext Markup Language* (HTML) é um conjunto *tags* que originou-se visando a exibição de textos digitais por intermédio de computadores na *WEB*, a partir desta premissa a linguagem será utilizada para adicionar tipagem através da suas *tags* identificando os elementos do texto que serão processados por um navegador ou browser para exibição na *WEB*.

Temos um exemplo de uma estrutura básica de HTML com suas respectivas *tags*, `<html></html>` indica o início e o fim do documento, `<head></head>` indica o cabeçalho do documento e ele não é visível na página, `<body></body>` indica todo o corpo do documento que é visível no navegador.

Figura 2 - Estrutura básica de HTML.



O conhecimento prévio do HTML é necessário tanto para melhoria na performance do SEO quanto para recuperação de dados através da prática de *WEB scraping*.

TRÄSEL (2014) define *WEB scraping* ou raspagem de dados como uma prática que transforma as páginas *WEB* em base de dados, no qual visa a automação do processo de obtenção dos dados das páginas que não disponibilizam esses dados por intermédio de dados estruturados tais como CSV, TXT ou XLS. A utilização dessa prática é implementada através da linguagem de programação, este artigo dará ênfase no uso da linguagem de programação python.

Segundo ROCHA (2006) linguagem de programação é um método padronizado que tem como objetivo instruir o computador a executar tarefas. Esse artigo utilizará a linguagem de programação python por apresentar uma boa compatibilidade com a *WEB*, bibliotecas de código úteis para o sistema tanto para a área de mineração de dados com *WEB scraping* quanto com a utilização de algoritmos de inteligência artificial. Para BORGES (2014) Python é uma linguagem de alto nível, que possui suporte ao paradigma de programação orientada a objeto, que possui tipagem dinâmica e forte e amplamente utilizado para desenvolvimento de sistema e *scripts* para automação de processos. Uma linguagem ideal para o uso do sistema na área de programação de linguagem natural PLN.

NUNES (2008) define PLN como o nome dado à área de pesquisa que analisa e propõe o desenvolvimento de sistemas e metodologias que visam a linguagem natural como prioridade. O conhecimento dessa área será útil para entender os motivos da implementação da *Natural Language Toolkit* (NLTK).

Para DE ALENCAR (2012) a biblioteca NLTK é um conjunto de ferramenta utilizada para implementação em conjunto da linguagem de programação python que possibilita a construção de *parsers*. O que no programa servirá como analisador de sintaxe da linguagem natural, no qual, a Inteligência Artificial utilizará para reconhecimento de padrões na linguagem.

DESENVOLVIMENTO DE FERRAMENTA SEO UTILIZANDO IA COM USO DE DADOS DOS BUSCADORES PARA MELHORIA DO RANQUEAMENTO NOS BUSCADORES: ESTUDO DE CASO DO SITE DHILUF TECH
Dhiego Costa Oliveira, Felipe de Oliveira C. de Sousa, Lucas Lira B. de Souza, Erwin Alexander Uhlmann

Segundo NIKOLOPOULOS (1997) a Inteligência Artificial é um campo multidisciplinar que tem como principal objetivo construir desenvolver um sistema capaz de tomar decisões e desempenhar ações similares a um ser humano. Essas tomadas de decisões utilizará um sistema de banco de dados para servir armazenar os dados que posteriormente servirá ao sistema como um todo.

Para MARA (2004) um sistema de gerenciamento de banco de dados é um sistema independente que será utilizado para armazenar arquivos no disco rígido para posteriormente utilizar esses arquivos.

DESENVOLVIMENTO

Para a confecção do sistema de produção de texto utilizando PLN com intuito de atuar no desenvolvimento de textos SEO, será utilizada na primeira fase a técnica de *WEB scrap* para extração de dados e textos que serão utilizados para o treinamento posterior do agente inteligente, sendo esse agente um atuador que irá classificar o texto extraído.

Com uma explicação sucinta pode-se dizer que *WEB scrap* é uma técnica de extração de dados automatizada voltada para páginas *WEB*. Um exemplo de uso da técnica pode ser descrito como, dada a necessidade de uma empresa de *e-commerce* saber como o serviço ofertado está sendo recebido pelo seus clientes, com a finalidade de melhorar o relacionamento entre a empresa e o cliente, foi contratado um profissional para aplicar essa pesquisa através do *feedback* nas redes sociais da empresa, no qual foi feita a extração de cada comentário referente a empresa que posteriormente foi tratado e quantificado para que um analista de dados pudesse apresentar. Concluindo que a técnica de *scrap* tem como principal objetivo a obtenção de dados da *WEB*.

Neste caso o intuito do sistema é trabalhar de maneira autônoma sem interferência externa, para que isso ocorra será utilizada a biblioteca NLTK para classificação das palavras de que trabalha de maneira similar a um compilador de linguagem de programação, essa similaridade é dada pois, será gerada uma tabela de símbolos esses contendo parâmetros como palavras, classificação, variação linguística, definição, sinônimos entre outros aspectos necessário para o vocábulo do português brasileiro.

Porém como o método adotado para a produção do sistema foi o de Corpus, onde este método atua de forma estatística e com grande volume de dados para melhor índice de acerto na produção de texto, será necessário o tratamento prévio de dados para que não haja contaminação na base dados por ocorrência de erros ortográficos ou erros gramaticais.

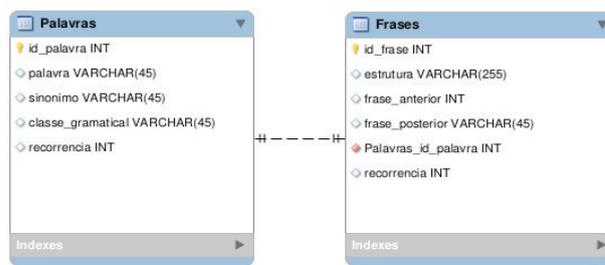
Com intuito de normalizar o texto extraído será utilizada uma ferramenta muito útil para se trabalhar com textos o *regex* ou *Regular Expression* onde sua

finalidade é a identificação, alteração e exclusão de cadeia de caracteres, através do *regex* o texto será normalizado para que só seja persistido na base de dados os textos de linguagem natural limpando qualquer tipo “ruído ou sujeira”, cujo possua uma cadeia de caracteres lógica, sejam *links*, *urls*, códigos indesejados.

Após a eliminação dos textos indesejados, será realizada a separação de frases para melhorar a velocidade de formulação dos textos através da recorrência das palavras em conjunto com o contexto, as frases separadas terão as suas estruturas analisadas pela NLTK e persistidas em um banco de dados para que o *software* possa se basear na produção um novo conteúdo de texto.

A Estrutura do banco de dados basicamente será desenhado da seguinte forma:

Figura 3 - MER do Banco de Dados atual.



A principal problemática do *software* é a lógica aplicada para a produção dos textos que não é genérica pois se baseia no método de corpus, onde esse *software* será treinado para lidar com apenas um segmento por vez, tendo assim uma estrutura de texto para cada segmento, assim apresentando um melhor comportamento quando usado de maneira específica para apenas um único segmento.

Após o treinamento da Inteligência Artificial para utilizar a estrutura textual mais recorrente, o texto será formulado inserindo as *keywords* desejadas transformando gerando assim um texto SEO.

RESULTADOS

Com implementação do sistema é esperado que a concorrência de conteúdo relevante para o *site* seja parelha a grandes competidores do mercado *WEB*, criando assim uma vantagem a partir da oportunidade de concorrer mesmo estando iniciando um novo *site*, tendo em vista que a criação para conteúdo de texto SEO é uma tarefa até o momento manual e lenta.

DISCUSSÃO

Algumas observações podem ser explanadas como a capacidade do *software* de alavancar o posicionamento do *site* que o estão utilizando, a proposta

do *software* não é de fato alavancar o posicionamento nos buscadores apesar de ser uma consequência, o enfoque do *software* está na automação para a produção de texto auxiliando o SEO, porém com o amadurecimento da ferramenta será possível que a produção de textos complexos a relevância do conteúdo seja eficiente a ponto de conseguir retirar de fato essa tarefa das mãos humanas.

Outra questão é compatibilidade do *software* com outros sistemas, *frameworks* e ferramenta de CMS, a princípio o *software* funcionará como uma Interface de Programação de Aplicativos API servindo através de requisições HTTP as mais variadas plataformas. São muitas as possibilidades e caminhos a serem seguidos, porém tudo dependerá de como o mercado se posicionará para que sim as problemáticas do *software* possam ser estudadas com mais abrangência.

CONCLUSÃO

Concluindo que com o auxílio da Inteligência Artificial utilizada no *software* será possível agilizar o processo do desenvolvimento de conteúdo de texto SEO que acarretará no impulsionamento do *site* nos mecanismos de buscas oferecendo a chance de competir em conteúdo com os competidores de ranque mais elevado.

REFERÊNCIAS

SILVA, Marcos José. et al. AC3AS-WEB: ambiente cooperativo de apoio à avaliação de aprendizagem significativa na WEB. In: Brazilian Symposium on Computers in Education (Simpósio Brasileiro de Informática na Educação-SBIE). 2002. p. 20-30.

SANTOS, João Batista Nascimento dos; KUNZ, Marinês Andrea. Estratégias de marketing em plataformas digitais. **Revista de Administração**, v. 12, n. 21, p. 73-92, 2014.

MENESES, Thiago Fonsêca; PIMENTEL FILHO, Carlos A. F.; ARAUJO, Rafael W. M. de **Guaatupi**: um ambiente para indexação e recuperação de imagens da WEB sem redundância visual.

SOUZA, Renato Rocha; ALVARENGA, Lídia. A WEB semântica e suas contribuições para a ciência da informação. **Ciência da Informação**, v. 33, n. 1, 2004.

TRÄSEL, Marcelo Ruschel. **Entrevistando planilhas**: estudo das crenças e do ethos de um grupo de profissionais de jornalismo guiado por dados no Brasil, 2014.

ROCHA, Rogério. **Utilização da robótica pedagógica no processo de ensino aprendizagem de programação de computadores**. Centro Federal e Educação Tecnológica de Minas Gerais, 2006.

BORGES, Luiz Eduardo. **Python para desenvolvedores**: aborda Python 3.3. Novatec Editora, 2014.

NUNES, Maria das Graças Volpe. **Processamento de línguas naturais**: para quê e para quem? ICMC-USP, 2008.

ALENCAR, Leonel Figueiredo de. Donatus: uma interface amigável para o estudo da sintaxe formal utilizando a biblioteca em Python do NLTK. **ALFA: Revista de Linguística**, v. 56, n. 2, 2012.

NIKOLOPOULOS, C. **Expert systems**: New York: Marcel Dekker, Inc., 1997.