

Unificação de Avaliações de Produtos em E-commerce: Uma Solução Computacional para Apoiar Decisões de Compra Online

Lívia Fernandes da Rocha.

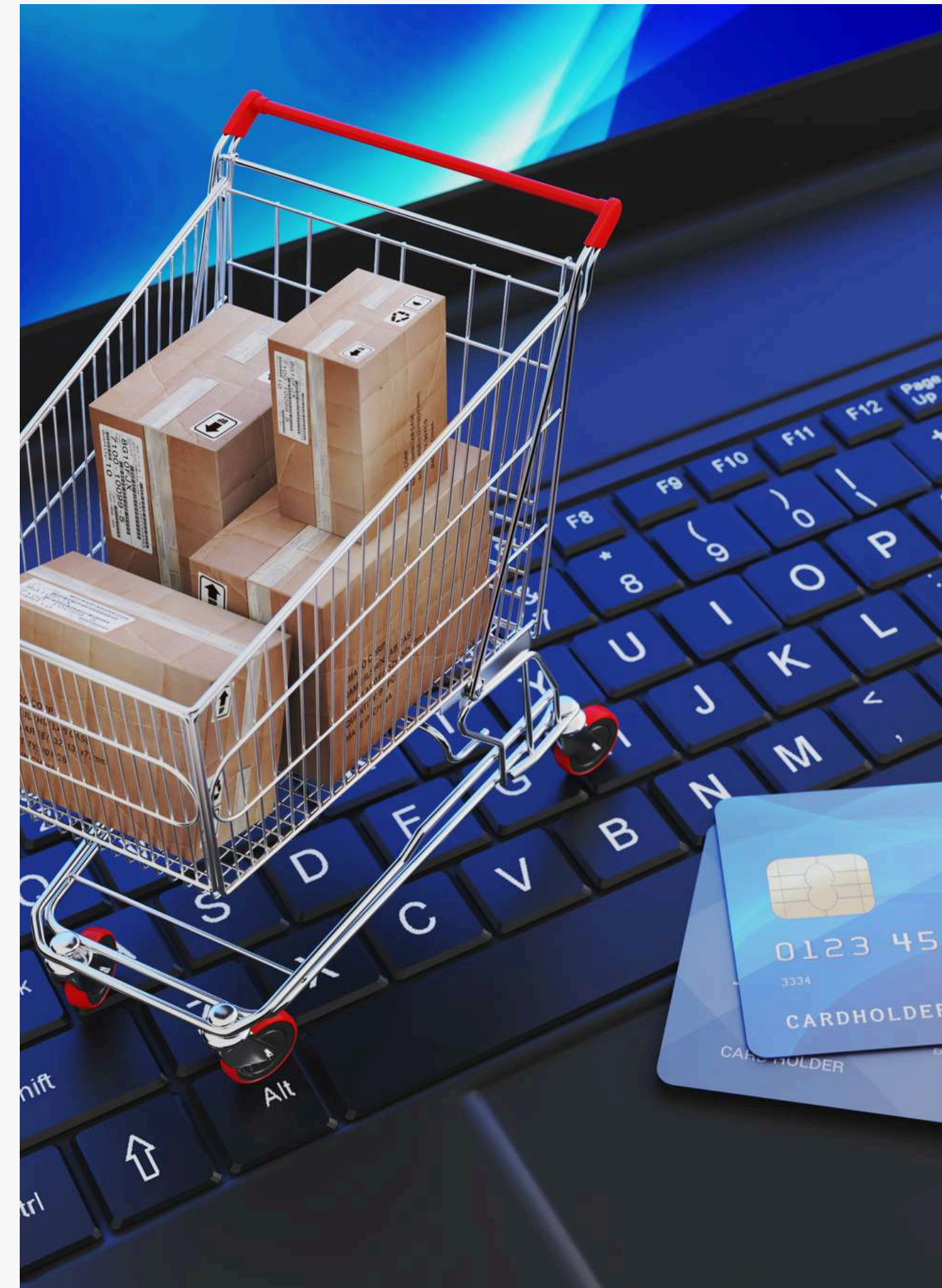
Orientador: Prof. Dr. Jorge Henrique Viana.

Sumário

1	Introdução	5	Análise e Caracterização dos Dados Coletados
2	Justificativa	6	Resultados Esperados
3	Objetivos	7	Referências
4	Metodologia		

Introdução

O comércio eletrônico tem se consolidado nas últimas décadas como um dos pilares do consumo contemporâneo, promovendo uma profunda transformação nas relações de compra e venda. A difusão da internet e a popularização dos dispositivos móveis impulsionaram a emergência de um ambiente de consumo digital, no qual milhões de transações são realizadas diariamente em escala global.



Justificativa

O projeto se justifica por sua capacidade de resolver a dispersão de avaliações de forma automatizada e eficiente. Ao centralizar dados de múltiplas fontes, a plataforma proposta otimiza o tempo do usuário e oferece uma perspectiva mais abrangente sobre a reputação e a qualidade dos produtos (*Filieri; Mcleay, 2014*).

- e-WOM (Electronic Word-of-Mouth)



Objetivos

01.	02.	03.	04.	05.	06.
Identificar e selecionar as principais plataformas de e-commerce brasileiras para a coleta de avaliações, considerando sua relevância e volume de dados.	Desenvolver e otimizar scripts de web scraping para a extração eficiente de dados dinâmicos, garantindo a adesão a boas práticas e políticas de uso dos websites.	Projetar e implementar um modelo de banco de dados para o armazenamento estruturado das avaliações e metadados dos produtos	Aplicar técnicas de Processamento de Linguagem Natural (PLN), como TF-IDF e similaridade de cossenos, para identificar e unificar produtos semanticamente semelhantes entre diferentes plataformas.	Desenvolver uma interface web intuitiva e responsiva para a consulta e a visualização das avaliações agregadas, com funcionalidades de busca e filtragem.	Implementar um mecanismo de ranqueamento que compare a performance de diferentes plataformas ou vendedores com base nas avaliações coletadas.

Metodologia

amazon

The Amazon logo consists of the word "amazon" in a bold, black, sans-serif font. Below the word is a thick orange curved arrow that starts under the 'a' and points towards the 'n', resembling a smile.

Metodologia



ASIN (Amazon Standard
Identification Number)

```
FUNÇÃO extrair_comentarios_amazon(html_page, asin, nome_produto, preco,
nota_geral):
    soup = BeautifulSoup(html_page, 'html.parser')
    comentarios_encontrados = []
    blocos =
    ENCONTRAR_TODOS_ELEMENTOS_COM_TAG_DIV_E_DATA_HOOK_REVIEW

    PARA CADA bloco EM blocos:
        texto = SELECIONAR_ELEMENTO_COM_DATA_HOOK_REVIEW_BODY
        nota =
        SELECIONAR_ELEMENTO_COM_DATA_HOOK_REVIEW_STAR_RATING
        data_tag = SELECIONAR_ELEMENTO_COM_DATA_HOOK_REVIEW_DATE

        pais, data_formatada =
        formatar_data_e_pais_amazon(data_tag.get_text())

        ADICIONAR comentario À comentarios_encontrados:
        "ASIN": asin
        "País": pais
        "Data Comentário": data_formatada
        "Nota": texto_da_nota
        "Comentário": texto_do_comentario
        "Nome": nome_produto
        ... outras informações do produto

    RETORNAR comentarios_encontrados
```


Metodologia



```
FUNÇÃO extrair_comentarios_mercadolivre(driver):
    comentarios = []

    TENTAR:
        CLICAR_NO_BOTAO_VER MAIS_OPINIOES_COM_DATA_TESTID_SEE_MOR
        E
        ESPERAR_POR_IFRAME_COM_ID_UI_PDP_IFRAME_REVIEWS
        MUDAR_DRIVER_PARA_IFRAME

        elemento_de_rolagem =
        ENCONTRAR_ELEMENTO_QUE_CONTEM_OS_COMENTARIOS

        last_height = 0
        LOOP infinito:
            EXECUTAR_SCRIPT_DE_ROLAGEM_DO_ELEMENTO
            ESPERAR_2_SEGUNDOS
            new_height =
            EXECUTAR_SCRIPT_PARA_VERIFICAR_TAMANHO_DA_PAGINA
            SE new_height == last_height:
                BREAK
            last_height = new_height

        comentarios_elementos = ENCONTRAR_TODOS_OS_ARTIGOS_DE_COMENTARIO

        PARA CADA comentario_el EM comentarios_elementos:
            texto = comentario_el.find_element(By.CSS_SELECTOR,
            "p.ui-review-capability...")
            nota = extrair_nota(comentario_el)
            data = extrair_data_comentario(comentario_el)

            ADICIONAR comentario À comentarios:
                "Comentário": texto_do_comentario
                "Nota": nota
                "Data Comentário": data

    FINALMENTE:
        MUDAR_DRIVER_PARA_CONTEUDO_PADRAO
        RETORNAR comentarios
```

Pseudocódigo da Extração de Comentários no Mercado Livre (scraperml.py)

Análise e Caracterização dos Dados Coletados



Quadro 1: Comparativo das Metodologias de Web Scraping.

Aspecto da Coleta	Amazon (Eletrônicos)	Mercado Livre (Eletrodomésticos)
Acesso	Requer autenticação por login, com cookies persistidos.	Acesso público, sem necessidade de login ou cookies.
Identificador Único	ASIN (Amazon Standard Identification Number).	ID numérico (ex: MLB15578949).
Extração de Produtos	Extração de ASINs da página de bestsellers.	Extração de IDs e links da página de mais vendidos.
Carregamento de Comentários	Rolagem da página para carregar conteúdo incremental.	Clicar no botão ("Mostrar todas as opiniões") e simular rolagem em um iframe.
Campos Coletados	Posição, ASIN, Nota Geral, Qtd. Avaliações, País, Data Comentário, Nome, Preço à vista, Nota, Comentário, Link.	Posição, ASIN (ID), Nota Geral, Qtd. Avaliações, Data Comentário, Nome, Preço à vista, Nota, Comentário, Link.

Análise e Caracterização dos Dados Coletados

Quadro 2 : Resumo geral dos produtos e avaliações na categoria “Eletrodomésticos”.

Plataforma	Total de Produtos	Total de Avaliações	Média de Avaliações por Produto	Média das Notas
Amazon	5	18.384	3.676,8	4,53
Mercado Livre	5	304.261	60.852,2	4,72

Resultados Esperados e Próximos Passos

- **Integração dos Dados**

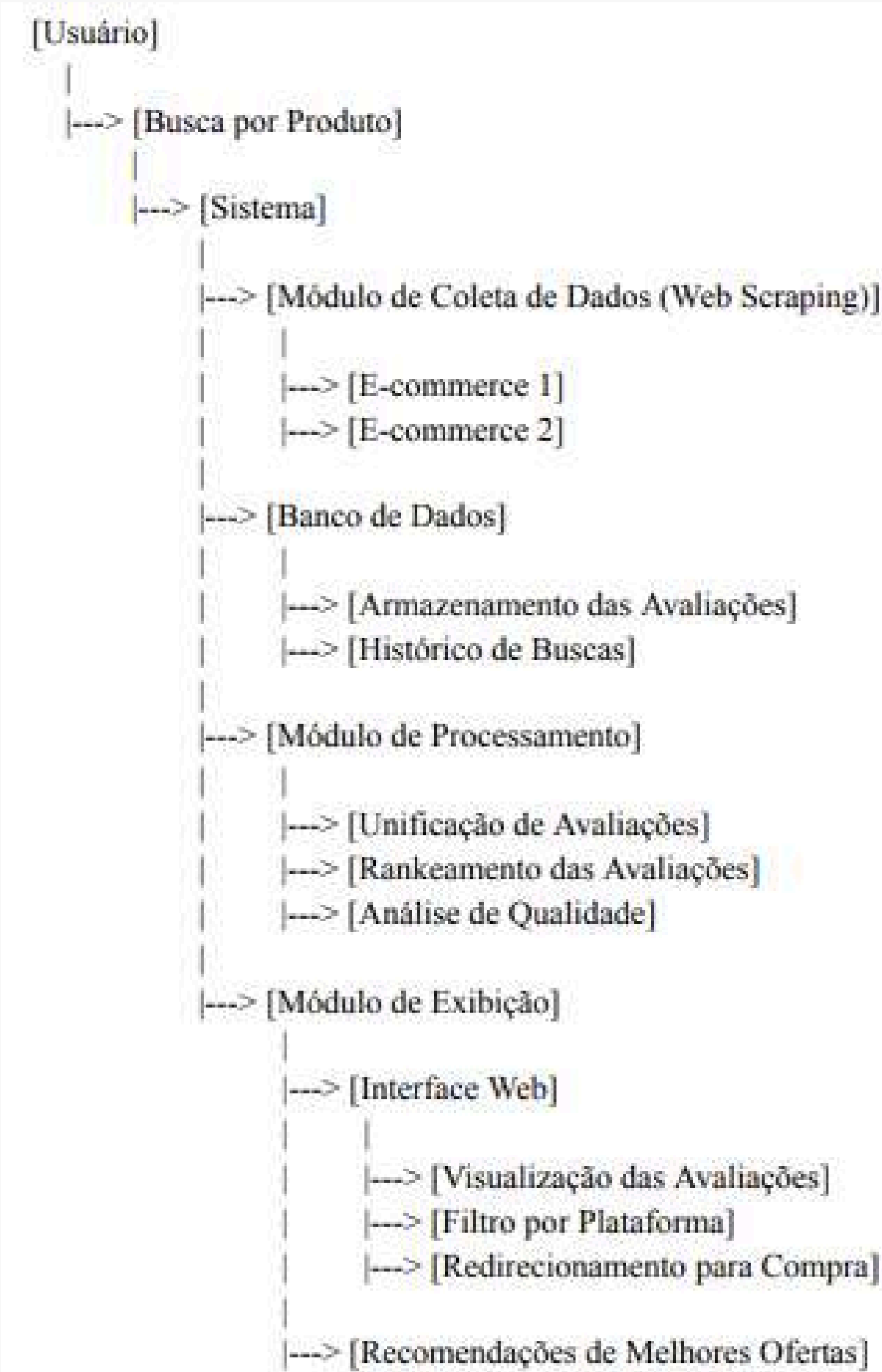
O modelo ***TF-IDF (Term Frequency-Inverse Document Frequency)*** será utilizado para converter os títulos dos produtos em vetores numéricos. Em seguida, a ***similaridade de cossenos (cosine similarity)*** será aplicada para calcular a proximidade entre esses vetores, permitindo a identificação e o agrupamento de produtos correspondentes com alta precisão (Manning; Raghavan; Schütze, 2008).

- **Implementação de um Protótipo Funcional**

Para cada modelo de produto encontrado, como um liquidificador específico, o usuário poderá clicar para acessar um dashboard detalhado. Esta tela individualizada é o núcleo da plataforma, onde os dados coletados e processados são apresentados de forma estratégica.

"Em qual loja este produto é mais bem avaliado pelos compradores?"

Mapa Conceitual do Sistema



OBRIGADA!