



ANÁLISE DE INTERESSE DE VISITANTES POR CATEGORIAS E KEYWORDS


CE062 - Mineração de Texto - 2017

Daniel Tyszka Jr / [github](#)

RELEMBRANDO...

 Descrição

 Características

 Comentários

▶ Detalhes

O novo Piacere Vetro com mesa em vidro temperado é o reflexo do design e sofisticação em sua cozinha. Praticidade e beleza refletindo todo o seu bom gosto. O Fogão Piacere Vetro oferece desempenho gourmet com as diversas potências de queimadores e detalhes exclusivos, como o acabamento na mesa e as grades com sistema de encaixe diferenciado, que garantem painéis mais estáveis e seguros. Tudo isso para deixar seu dia a dia muito mais prático para que você brilhe com suas melhores receitas.

id	Visitante1	Visitante2
Categoria	Fogões	Fogões
Keywords	vidro temperado, desempenho gourmet, painéis mais estáveis e seguros	grades individuais, acendimento automático, forno esmaltado

OBTE[~]NÇÃO DE DADOS

- Páginas: scrapping (40k total, 5k produtos)
- Visitantes: simulação (15k)

EXTRAÇÃO E CLASSIFICAÇÃO

Tipo de Página e Categorias: variável "dataLayer"

```
<title>Umificador - RicardoEletro.com</title>
<link rel="icon" href="http://images.maquinadevendas.com.br/site/re/favicon.ico" type="image/x-icon" />
<meta name="google-site-verification" content="hgorFwvgMHZXXUmeGTufBfhJDEM7AjkfeRHlN3ND8rc" />
</head>
<body>
<script>
var dataLayer = [{
  "pageDepartment": "Beleza e Saúde",
  "pageCategory": "Cuide de sua Saúde",
  "pageSubCategory": "Umificador",
  "pageDepartmentID": 1,
  "pageCategoryID": 40,
  "pageSubCategoryID": 66,
  "pageTitle": "Ricardo Eletro aqui tem Dono, na luta pelo menor preço!",
  "categoryFilters": [],
  "productList": [
    {
      "id": "82745",
      "sku": "94421",
      "name": "Umificador de Ar Ultrassônico 2,2L G-Tech - Névoa Finíssima, Compacto, Leve e Silencioso, Desliga Automático, Bivolt - Allergy Free Baby",
      "price": "54.9"
    },
    {
      "id": "193464",
      "sku": "235218",
      "name": "Umificador Ultrassônico 3,5L Bivolt - Ventisol",
      "price": "69.9"
    },
    {
      "id": "559733",
      "sku": "640228",
      "name": "Umificador de Ar Ultrassônico 2,5L Wiso - Melhora a Qualidade do Ar, Dispersão Ultrafina, Silencioso, Regulável WU125",
      "price": "239.9"
    }
  ]
}]
</script>
<div id="Busca-Cemporcento" class="content-full">
```

EXTRAÇÃO E CLASSIFICAÇÃO

Descrição do Produto: html

```
<div class="conteudo-abas" id="aba-descricao">
<h3 class="subtitulo-sessao" id="DetalhesAncora">
<span class="indicador">
</span>Detalhes
</h3>
<div class="descricao-produto">
<p class="MsoNormal" style="BACKGROUND: white; LINE-HEIGHT: 9pt">
</p>
<p class="MsoNormal">O umidificador G-TECH Allergy Free Filter Dual 3 L, é ideal para utilização em ambientes muito secos ou
que façam uso prolongado de Ar-Condicionado, como escritórios. Sua ampla capacidade do reservatório permite uma grande
autonomia de funcionamento ( até 23 horas), abrangendo ambientes de até 30m2. Possui o exclusivo saída dupla de névoa,
aumentando a performance e o alcance da umidificação e com seu sistema de desligamento automático na falta de água no
reservatório, garante a segurança e evitando acidentes.
<br />* Saída dupla de névoa, melhor performance na umidificação do ambiente.
<br />* Gerador ultra sônico de Alta Frequência que produz finíssima névoa.
<br />* Desligamento automático quando não há água no reservatório.
<br />* Super silencioso quando está operando.
<br />* Baixíssimo consumo de energia.
<br />* Reservatório para 3L.
<br />* Bivolt.
<br />* Até 10 horas de autonomia em potência máxima.
<br />* 1 Ano de Garantia.
</p>
</div>
</div>
<div class="conteudo-abas" id="aba-caracteristicas">
```

TRATAMENTO DE KEYWORDS

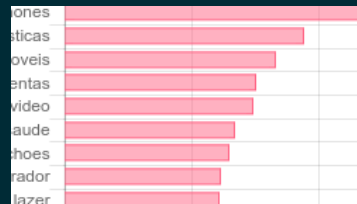


Padrões?

[substantivo]	[substantivo][advérbio]	[substantivo][adjetivo]	[substantivo][preposição][substantivo]
qualidade	elegância à	sistema ciclônico	programas de lavagem
produto	utilização quando	cabo elétrico	consumo de energia
embalagem	noite bem	aspirador smart	função com acionamento
caixas	bambu ecologicamente	reservatório prático	chapa de aço
produto	fato bem	cabo elétrico	kit de acessórios

CRUZAMENTOS

Dados de visitas x Dados sobre Páginas



CONSIDERAÇÕES FINAIS

- ML para Categorizador e Extrator de Dados (tipo de página, categorias/departamentos, descrições de produtos)
- OCR para descrições de produtos em imagens
- POS Tagger em português deixa a desejar
- Extração de keywords por padrão POS requer atenção dedicada
- Falta de "rank de importância" nas keywords

ANÁLISE DE INTERESSE DE VISITANTES POR CATEGORIAS E KEYWORDS

CE062 - Mineração de Texto - 2017

Daniel Tyszka Jr / [github](#)