UNIVERSIDADE FEDERAL DE SANTA CATARINA – UFSC
FACULDADE DE CIÊNCIA DA COMPUTAÇÃO PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

JOSÉ HENRIQUE CALENZO COSTA

FILTERED-PAGE RANKING: UMA ABORDAGEM PARA RANQUEAMENTO DE DOCUMENTOS HTML PREVIAMENTE FILTRADOS

ORIENTADORA
PROFA. DRA. CARINA F. DORNELES

FLORIANÓPOLIS
2016

Ficha catalográfica

Costa, José Henrique Calenzo
FILTERED-PAGE RANKING: UMA ABORDAGEM PARA RANQUEAMENTO DE DOCUMENTOS HTML PREVIAMENTE FILTRADOS / José Henrique
Calenzo Costa ; orientadora, Carina Friedrich Dorneles -
Florianópolis, SC, 2016.
110 p.
Dissertação (mestrado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação Multidisciplinar em Saúde.
Inclui referências
1. Saúde. 2. Ranking de Páginas HTML. 3. Segmentação de Páginas HTML. 4. Remoção de Conteúdo Irrelevante em Páginas HTML. I. Dorneles, Carina Friedrich. II. Universidade Federal de Santa Catarina. Programa de Pós-Graduação Multidisciplinar em Saúde. III. Título.

José Henrique Calenzo Costa

FILTERED-PAGE RANKING: UMA ABORDAGEM PARA RANQUEAMENTO DE DOCUMENTOS HTML PREVIAMENTE FILTRADOS

Dissertação submetida ao Programa de Pós-Graduação em Ciência da Computação, da Universidade Federal de Santa Catarina, para a obtenção do Grau de Mestre em Ciência da Computação.

Orientadora: Profa. Dra. Carina F. Dorneles

Florianópolis

2016

José Henrique Calenzo Costa

FILTERED-PAGE RANKING: UMA ABORDAGEM PARA RANQUEAMENTO DE DOCUMENTOS HTML PREVIAMENTE FILTRADOS

Esta dissertação foi julgada adequada para a obtenção do título de `mestre_, e aprovada em sua forma final pelo Programa de Pós-Graduação em Ciência da Computação.

Florianópolis, 10 de março de 2016.

_____

Profa. Dra. Carina Friedrich Dorneles
Coordenadora do Programa

Banca Examinadora:

_____

Profa. Dra. Carina Friederich Dorneles
Presidente (orientadora)
Universidade Federal de Santa Catarina

_____

Profa. Dra. Renata de Matos Galante
Universidade Federal do Rio Grande do Sul

_____

Prof. Dr. Roberto Willrich
Universidade Federal de Santa Catarina

_____

Prof. Dr. Ronaldo dos Santos Mello
Universidade Federal de Santa Catarina

Este trabalho é dedicado à minha orientadora, à minha filha, à minha esposa e aos meus queridos amigos.

AGRADECIMENTOS

Agradeço à minha orientadora, à minha esposa e à minha filha.

“Não pense que é capaz. Saiba que é.”Matrix

RESUMO

Algoritmos de ranking de páginas Web podem ser criados usando técnicas baseadas em elementos estruturais da página Web, em segmentação da página ou na busca personalizada. Esta pesquisa aborda um método de ranking de documentos previamente filtrados, que segmenta a página Web em blocos de três categorias para delas eliminar conteúdo irrelevante. O método de ranking proposto, chamado Filtered-Page Ranking (FPR), consta de duas etapas principais: (i) segmentação da página web e eliminação de conteúdo irrelevante e (ii) ranking de páginas Web. O foco da extração de conteúdo irrelevante é eliminar conteúdos não relacionados à consulta do usuário, através do algoritmo proposto Query-Based Blocks Mining (QBM), para que o ranking considere somente conteúdo relevante. O foco da etapa de ranking é calcular quão relevante cada página Web é para determinada consulta, usando critérios considerados em estudos de recuperação da informação. Com a presente pesquisa pretende-se demonstrar que o QBM extrai eficientemente o conteúdo irrelevante e que os critérios utilizados para calcular quão próximo uma página Web é da consulta são relevantes, produzindo uma média de resultados de ranking de páginas Web de qualidade melhor que a do clássico modelo vetorial.

Palavras-chave: Remoção de ruídos em páginas HTML. Ranking. Extração automática de conteúdos WEB.

ABSTRACT

Web page ranking algorithms can be created using content-based, structure-based or user search-based techniques. This research addresses an user search-based approach applied over previously filtered documents ranking, which relies in a segmentation process to extract irrelevante content from documents before ranking. The process splits the document into three categories of blocks in order to fragment the document and eliminate irrelevante content. The ranking method, called Page Filtered Ranking, has two main steps: (i) irrelevante content extraction; and (ii) document ranking. The focus of the extraction step is to eliminate irrelevante content from the document, by means of the Query-Based Blocks Mining algorithm, creating a tree that is evaluated in the ranking process. During the ranking step, the focus is to calculate the relevance of each document for a given query, using criteria that give importance to specific parts of the document and to the highlighted features of some HTML elements. Our proposal is compared to two baselines: the classic vectorial model, and the CETR noise removal algorithm, and the results demonstrate that our irrelevante content removal algorithm improves the results and our relevance criteria are relevant to the process.

Keywords: HTML noise removal. Page segmentation. Ranking. Web content automatic extraction.

LISTA DE FIGURAS

Figura 1 – Histograma CETR 30
Figura 2 - Árvore DOM antes QBM 43
Figura 3 - Árvore DOM após QBM 44
Figura 4 - Algoritmo FPR 50
Figura 5 - Precisão P@10, P@15, P@20 60
Figura 6 - F-Measure 20 primeiras posições 61
Figura 7 - Curva Precisão x Revocação 62

LISTA DE TABELAS

Tabela 1 – Termos Fórmula Ranking Lucene 25
Tabela 2: Algoritmos de Extração de dados de páginas Web 31
Tabela 3 - Algoritmos de Ranking na WEB 36
Tabela 4 - Consultas realizadas para os experimentos 55
Tabela 5 - Revocação e Precisão - QBM x CETR..... 57

LISTA DE ABREVIATURAS E SIGLAS

CETR	Content Extration via Tag Ratios
DOM	Document Object Model
DSE	Data-Rich Section Extraction
FPR	Filtered Page Ranking
HITS	Hipertext Induced Topic Search
HTML	HiperText Markup Language
IDF	Inverse Document Frequency
MDR	Overall Algorithm
QBM	Query-Based Blocks Mining
QTDPDF	Total de Palavras da DOM filtrada
QTDPRP	Quantidade de Páginas Recuperadas na Consulta
QTDPUR	Quantidade de Páginas Úteis Recuperadas na Consulta
QTDUDF	Quantidade de Palavras Úteis na DOM Filtrada
QTDUDO	Quantidade de Palavras Úteis na DOM Original
TF	Term Frequency
TTR	Text to Tag Ratios
WLR	Words/Leafs Ratio

SUMÁRIO

1 INTRODUÇÃO.....17
1.1 PROBLEMA.....19
1.2 SOLUÇÃO.....20
1.3 OBJETIVO GERAL.....20
1.3.1 Objetivos Específicos.....20

2 TRABALHOS RELACIONADOS.....23
2.1 BASE CONCEITUAL.....23
2.1.1 Árvore DOM.....23
2.1.2 Lucene.....23
2.2 EXTRAÇÃO DE RUÍDOS EM PÁGINAS HTML.....25
2.2.1 Abordagens de extração de dados ou segmentação da página WEB.....25
2.2.2 Análise Comparativa.....31
2.3 RANKING DE DOCUMENTOS HTML.....32
2.2.3 Análise Comparativa.....35

3 FILTERED-PAGE RANK.....39
3.1 VISÃO GERAL.....39
3.2 QUERY-BASED BLOCKS MINING.....40
3.2.1 Categorização dos blocos.....41
3.2.2 Criação da árvore filtrada.....42
3.3 RELEVÂNCIA DE UMA PÁGINA HTML.....44
3.4 FUNÇÃO DE RANKING E O ALGORITMO FPR.....47

4 AVALIAÇÃO EXPERIMENTAL.....53
4.1 CONFIGURAÇÃO DAS CONSULTAS E DO CONJUNTO DE DOCUMENTOS.....54
4.2 METODOLOGIA E MÉTRICAS DE AVALIAÇÃO.....55
4.3 ANÁLISE DO QBM.....56
4.4 ANÁLISE DO FPR.....60
CONCLUSÕES E TRABALHOS FUTUROS.....63
REFERÊNCIAS.....65
ANEXOS.....69

INTRODUÇÃO

O processo de ranking de documentos surge em muitas aplicações, tais como máquinas de busca (LANGVILLE and MEYE, 2011; BRIN and PAGE, 2012; SU et al. 2014), sistemas de recomendação (KARATZOGLOU et al. 2013; LERCHE and JANNACH, 2014; SONG et al. 2011; BALOG and RAMAMPIARO 2013) e classificação de documentos (BERARDI et al. 2015; FANG et al. 2007; LI et al. 2011; CHEN et al. 2014; ZHU et al. 2013), entre outros.

As abordagens propostas variam e usualmente definem parâmetros relevantes para o ranking. O processo de ranking de documentos tem sido tratado, tradicionalmente, como um problema de correspondência (matching) entre uma consulta e um conjunto de documentos. Neste contexto, um desafio comum é encontrar uma maneira de selecionar documentos representativos para uma consulta específica e explorar novos modelos de classificação que produzam resultados precisos para cada consulta.

A explosão de dados publicados na Web traz uma grande quantidade de conteúdo interessante e útil, mas, ao mesmo tempo, uma grande quantidade de dados considerados ruídos, como banners, slogans, imagens comerciais, menus, rodapés, etc., considerados irrelevantes, além de dificultarem o processo de classificação usando os algoritmos mais clássicos que tratam toda informação apresentada nos documentos de maneira uniforme, sem desconsiderar áreas com dados inúteis para a busca efetuada. Espera-se um melhor resultado de ranqueamento de documentos com a remoção de conteúdos irrelevantes.

A comunidade de recuperação da informação tem se preocupado com este fato. Antes da etapa de ranking, alguns estudos foram desenvolvidos a fim de excluir o máximo de ruídos de páginas Web (LIU 2011; BURGET and RUDOLFOVA 2009; INSA et al. 2013; VELLOSO and DORNELES 2013; WENINGER et al. 2010). Neste contexto, diferentes abordagens são desenvolvidas como técnicas baseadas na análise textual de documentos (WENINGER et al. 2010), em informações visuais (BURGET and RUDOLFOVA, 2009) e em

padrões de elementos (VELLOSO and DORNELES 2013; CRESCENZI et al. 2001) dentre outros. Ocorre, também, que mais de uma técnica pode ser considerada (LIU, 2011). A maioria das técnicas de extração de ruídos encontradas na literatura não considera a consulta do usuário.

O conteúdo dos documentos ranqueados é crucial para a qualidade de um ranking. Esperam-se melhores resultados de ranking com a limpeza do conteúdo, ou seja, com a remoção de ruídos. Em geral, apenas a região principal de uma página é de interesse, ou com informação útil, sendo outras regiões do documento consideradas ruídos e portanto, não relevantes.

Não existe uma forma única de realizar o ranking de um conjunto de documentos. É possível, por exemplo, utilizar agregação de ranking (JOACHIMS, 2002), onde o resultado de um ranking é utilizado como peso para a geração de um novo ranking. Agregação de ranking é a combinação de diferentes ordens de rankings, de modo a obter uma "melhor" ordenação (DWORK et al., 2001). A recuperação da informação na Web na forma de rankings individuais é abundante e é geralmente baseada em métodos de recuperação da informação, em algoritmos baseados na análise de links e outros, usados para computar a relevância da pesquisa no documento. O resultado pode variar muito de acordo com o critério do ranking (ADALI et al., 2006).

O presente trabalho apresenta um método de ranking chamado **Filtered-Page Ranking** (FPR) que realize o ranqueamento de uma base de documentos HTML levando-se em consideração a proximidade das páginas Web com a consulta realizada pelo usuário através de duas etapas principais:

(i) extração de conteúdo irrelevante com fragmentação de página; e
(ii) (ii) ranking dos documentos.

O objetivo do processamento executado na etapa de extração é eliminar do documento original conteúdos não relacionados à consulta efetuada, através do algoritmo proposto *Query-Based Blocks Mining* (QBM), gerando um documento filtrado, que é avaliado na etapa de ranking. A remoção de conteúdo irrelevante de páginas Web utiliza-se

da segmentação, em blocos, de uma página Web, para posterior ranqueamento destes documentos. A segmentação realizada é baseada nos termos de busca do usuário e nas características de delimitação de segmentos de determinadas tags HTML; os documentos são segmentados em *blocos-segmento*, *blocos-destaque* e *blocos-descarte*, sendo excluídos os blocos considerados irrelevantes. Durante a etapa de ranking, são utilizados critérios de relevância para quantificar quão próximo o conteúdo de uma página é dos termos da busca. O ranking foca na definição da relevância das páginas HTML para uma dada consulta.

Para avaliar a proposta, experimentos são realizados sobre um repositório de páginas HTML. Os resultados são comparados nos aspectos de ranking e extração com dois baselines: o algoritmo de ranking do modelo vetorial (SALTON and BUCKLEY, 1988), através da implementação no Lucene, e o algoritmo de remoção de ruídos chamado CETR (WENINGER et al., 2010). Os experimentos demonstram que a etapa de extração de conteúdo irrelevante por meio da abordagem proposta melhora os resultados da busca e que os critérios utilizados para cálculo da relevância das páginas Web são relevantes no processo.

O Lucene (HATCHER et al., 2004) será utilizado como parâmetro por ser uma ferramenta (biblioteca de classes) de código aberto, bastante utilizada como núcleo de várias ferramentas de buscas locais, seja de páginas Web, seja de qualquer outro formato de documento do qual se possa extrair texto. Do mesmo modo, o algoritmo CETR, que apresenta bons resultados na extração automática de ruídos e possui implementação em código aberto, disponível em linguagem java, a mesma utilizada na elaboração da ferramenta de ranking proposta.

1.1 PROBLEMA

Os estudos de ranking de documentos e extração de conteúdo útil são muito utilizados em pesquisas de datamining. Poucos estudos mostram rankings de documentos com extração de conteúdo com a finalidade de que o ranking despreze conteúdo irrelevante. A união de

algoritmos de extração e ranking é encontrada no trabalho de Wang and Lochovsky (2002), que unem extração automática da região principal de páginas HTML e o algoritmo de ranking HITS. Também há trabalhos que utilizam segmentação de página no processo de ranking (SANCHES, 2013; CALLAN 1994; FERNANDES et al., 2007; CAI et al., 2004). Não encontramos algoritmos de ranking que avaliem quão relevante seja uma página Web para uma consulta efetuada e que leve em consideração aspectos da linguagem HTML (como palavras em destaque, no título, em links, etc.), com implementação, ou, pelo menos, critérios, pesos e fórmula de ranking disponíveis para comparação.

1.2 SOLUÇÃO

Propor um método de ranking que retire conteúdos irrelevantes de páginas HTML e ranqueie o conteúdo relacionado à busca efetuada pelo usuário, focando o ranking em uma página cujo conteúdo (links, textos, tabelas...) não relacionado à busca tenha sido removido.

1.3 OBJETIVO GERAL

O principal objetivo do presente trabalho é definir uma abordagem de construção de ranking de documentos HTML, a fim de que este ranking seja realizado somente sobre conteúdo útil para a busca do usuário, o que será feito unindo mineração de dados e critérios vistos em estudos de indexação e rankings na Web para concepção do método proposto ***Filtered-Page Ranking*.**

1.3.1 Objetivos Específicos

Objetivamos, especificamente, com o presente trabalho, propor:

- um método de ranking de páginas Web com extração de ruídos que possa ser aplicado em um repositório de páginas HTML, com índices de precisão e revocação melhores que os do Lucene;
- um método de extração de dados em páginas Web que elimine conteúdo não relacionado às palavras-chaves pesquisadas, permitindo concentrar o processamento de aplicações de indexação

e ranking na parte mais relevante da página.

No presente trabalho, é proposto um método novo de extração de conteúdo útil, com base nas palavras-chave da consulta e avaliamos a revocação e a precisão desta extração da região relevante, além de estabelecer os critérios que serão avaliados para a definição do ranking e seus pesos. Abordaremos trabalhos relacionados à extração automática de ruídos em páginas HTML na seção 2.2, ranking de páginas HTML na seção 2.3, em seguida, apresentaremos a proposta deste trabalho na seção 3, que une extração automática de conteúdo em páginas Web com ranking; experimentos na seção 4 e conclusão e possíveis trabalhos futuros na seção 5.

2 TRABALHOS RELACIONADOS

Este capítulo apresenta trabalhos que abordam técnicas de extração de ruídos em páginas HTML e métodos de ranking de páginas HTML, temática pertinente à tratada nesta pesquisa.

2.1 BASE CONCEITUAL

Antes de introduzirmos métodos encontrados na literatura de extração de conteúdo em páginas Web e de ranking de documentos, é importante definir o conceito de árvore DOM utilizado em muitos algoritmos de extração além de detalharmos a ferramenta Lucene que utiliza o clássico modelo vetorial para determinar quão próximo um documento é de uma consulta efetuada.

2.1.1 Árvore DOM

Antes de entrar nos métodos de extração de conteúdo em páginas WEB, apresentamos o conceito de árvore DOM, utilizado em muitos algoritmos de mineração de documentos HTML.

Árvore DOM é a representação de um documento HTML em estrutura de árvore, também conhecida como "árvore de tags". É construída a partir do código HTML de uma página, com a utilização de um *parser*. Como a tarefa de *parsing* de um código-fonte HTML pode tornar-se extremamente complicada devido à grande quantidade de códigos mal escritos que se encontram na WEB, é comum a utilização de um renderizador HTML para auxiliar nesta tarefa. O fato de existir grande quantidade de códigos mal escritos na WEB se deve à própria linguagem HTML, por possuir "*forgiving syntax*", o que permite que o código seja apresentado corretamente em um *browser*, mesmo quando desrespeita a sintaxe da linguagem.

2.1.2 Lucene

Lucene é uma biblioteca escalável de recuperação da

informação de alta performance que permite adicionar capacidades de pesquisa textual na aplicação. É um projeto de código aberto, maduro, implementado em java e do Apache Software Foundation (HATCHER et al., 2004).

Hatcher et al. (2004) esclarecem que Lucene não é uma ferramenta de pesquisa completa ou um motor de busca na WEB, mas uma biblioteca que se preocupa com indexação e pesquisa de qualquer documento do qual se possa extrair texto, como arquivos .html, .pdf, .doc, etc. As principais tarefas do Lucene são análise de documento, indexação e pesquisa. Para que o Lucene possa indexar um documento, é necessário informar qual indexador será utilizado pela classe responsável por gerar os índices. O Standard Analyzer é o mais sofisticado. O Lucene não torna o resultado da indexação visível ao usuário, sendo função do QueryParser analisar que documentos se relacionam melhor com a busca efetuada. O QueryParser do Lucene permite expressões sofisticadas na busca como *"presidente obama"* e *+harvard +professor*, buscando documentos que contenham a frase *presidente obama* (no primeiro caso) e documentos que contenham *harvard* e *professor* (no segundo caso).

Sempre que um documento combina com uma pesquisa, o Lucene computa um escore, número de relevância. Escores altos refletem mais similaridade do documento com a busca efetuada.

Existem três modelos teóricos de pesquisa textual:

i. *modelo booleano;*
ii. *modelo espaço-vetorial* e
iii. *modelo probabilístico*.

O Lucene trabalha tanto com o espaço vetorial quanto com o booleano. No booleano, o resultado indica se o documento combina ou não com a consulta, sem escore de relevância. Na pesquisa do trabalho, utilizamos o modelo espaço-vetorial, conhecido por vector space model, por o considerarmos o mais indicado para consultas com grandes quantidades de documentos, nas quais poderemos querer somente os top *N* documentos mais relevantes. No vector space model, a relevância é a medida da distância entre a pesquisa e o documento.

Molková (2011) calcula a fórmula de similaridade do Lucene

utilizando os parâmetros *TF-IDF* (*TF* = term frequency/frequência do termo, *IDF* = inverse document frequency/frequência do documento inversa) para cada termo da fórmula:

$$score(q,d) = coord(q,d) * queryNorm(q) * \sum [tf(t \in d) * idf(t)^2 * boost(t.field \in d) * lengthNorm(t.field \in d)]$$

Tabela1 – Termos Fórmula Ranking Lucene

Fórmula	**Explicação**
Q	Pesquisa, busca efetuada.
D	Documento.
$tf(t \in d)$	Número de vezes em que o termo aparece no documento.
$idf(t)$	Número de documentos que contêm o termo/número de documentos.
$coord(q,d)$	Quantos termos da pesquisa aparecem no documento.
$queryNorm(q)$	Normalizar o escore das pesquisas, dividindo pelo maior resultado. 1/maior escore.
$boost(t.field \in d)$	O custo da indexação do termo no documento, especificado durante o processo de indexação.
$lengthNorm(t.field \in d)$	Fator normalizador de termos curtos e de grande custo de indexação.

Fonte: Molková, 2011.

2.2 EXTRAÇÃO DE RUÍDOS EM PÁGINAS HTML

2.2.1 Abordagens de extração de dados ou segmentação da página WEB

Cerca de 40% a 50% do conteúdo apresentado em uma página

pode ser considerado irrelevante (INSA, 2013). A extração automática do conteúdo principal de páginas HTML permite que ferramentas de indexação sejam mais precisas eliminando conteúdos irrelevantes. Essas técnicas variam em escala e automatização. Para algumas, é necessário conhecimento prévio de uma página do site que servirá de modelo para extração de conteúdo, como é o caso do RoadRunner (CRESCENZI, 2001). Outros métodos, como o MDR (LIU, 2011), percorrem a árvore DOM em busca da localização de registros de dados. Há também outras técnicas de extração automática de conteúdo HTML, que levam em consideração a razão de palavras por nó para inferir a região principal e descartar blocos de dados irrelevantes (INSA, 2013).

Toda técnica de extração automática de conteúdo HTML é uma forma de mecanizar um trabalho que poderia ser feito manualmente se o conjunto de páginas fosse pequeno. Na prática, porém, isso não ocorre devido ao grande volume de dados e páginas existentes nas bases alcançadas pelas ferramentas de pesquisa. Quanto mais automatizado é o método, quanto menos intervenção humana, menor a precisão. Daí a importância do estudo da extração automática, ou semiautomática, de dados na Web para que se aumentem a precisão e a qualidade da extração e do resultado das ferramentas de indexação e ranking.

Técnicas de extração de ruídos podem ser construídas sob várias perspectivas. Neste trabalho, são propostas três categorias distintas:

(i) **técnicas baseadas na análise de texto presente nas tags ou linhas do documento:** nessas técnicas, a ideia principal é encontrar o subconjunto de tags ou região do documento HTML que possui a maior parte do conteúdo textual (INSA et al.. 2013; WENINGER et al., 2010);

(ii) **técnicas baseadas em informações visuais:** consideram aspectos de apresentação do conteúdo como diferenças de estilo, cor ou tamanho da fonte, para realizar a segmentação (CAI et al. 2003; BURGET and RUDOLFOVA; 2009);

(iii) **técnicas de extração baseadas em padrões de tags ou de apresentação de registros HTML:** procuram por similaridades entre páginas e um modelo ou por dados que apresentam

determinado padrão (registros simples ou aninhados) para extração da região de dados (CRESCENZI et al., 2001; WANG and LOCHOVSKY, 2002; LIU, 2011; VELLOSO and DORNELES, 2013).

O algoritmo ***RoadRunner*** (CRESCENZI et al., 2001) é utilizado para extração de dados de múltiplas páginas. Dado um conjunto de páginas similares, geralmente de um mesmo *site*, contendo um ou mais registros, o algoritmo compara as páginas para encontrar similaridades e diferenças entre elas, gerando uma expressão regular durante o processo. A cada comparação, a expressão regular é refinada/generalizada, resolvendo as diferenças encontradas, da seguinte maneira:

i. diferenças entre texto indicam campos de dados;
ii. diferenças entre *tags* indicam itens opcionais ou listas.

Após o término das comparações, a expressão regular é utilizada para extração de dados de outras páginas. A complexidade do algoritmo é exponencial em relação ao tamanho do código fonte HTML e não é eficiente para extrair registros em páginas com layouts com pouca semelhança, pela difícil determinação de um padrão.

O ***MDR*** (LIU, 2011) é utilizado na extração de registros simples de documentos HTML. O algoritmo percorre o DOM, recursivamente, em profundidade, no sentido da raiz para as folhas (*top-down*), procurando por padrões na árvore. Em cada nível da árvore, os filhos de cada nó (e as subárvores) são comparados entre si para verificar se existe alguma semelhança entre eles. As comparações são realizadas combinando os nós em grupos de 1 até *k* nós. Para evitar comparar todas as combinações possíveis de nós (o que seria proibitivo, devido à complexidade computacional), o algoritmo faz duas suposições a respeito dos registros que podem ser encontrados:

i. um grupo de registros de dados é apresentado em uma região contígua à página, utilizando estruturas HTML similares;
ii. uma lista de registros em uma região é formada por subárvores, no mesmo nível, filhas de um mesmo nó.

De acordo com essas possibilidades, é necessário realizar apenas as comparações com grupos de nós adjacentes. Sempre que o algoritmo inicia as comparações no nó seguinte, não é necessário realizar as comparações anteriores. Esta forma de comparar os nós reduz consideravelmente a complexidade computacional com relação à alternativa de comparar todas as combinações possíveis de nós. Para a comparação das subárvores, pode-se utilizar o algoritmo Simple Tree Matching (STM) (LIU 2011) para medir o grau de semelhança entre duas árvores. O MDR não necessita de um template ou realizar comparações entre páginas WEB para encontrar um padrão entre elas, sendo voltado à extração de registros de dados semiestruturados.

O ***TPS*** (VELLOSO and DORNELES, 2013), similarmente ao MDR, é um processo automático que não depende de páginas modelo para extração de registro de dados semiestruturados e leva em consideração a *tag-path* para definir regiões. Como a *tag-path* é o caminho do topo até qualquer nó na árvore DOM, sendo duas regiões da árvore DOM com *tag-paths* e estilos iguais, consideradas uma mesma região de registros de dados, este método parte da premissa de que regiões com dados semiestruturados com maior densidade de registros constituam a região principal da página WEB, sendo extraídas as regiões de dados cujos tag-paths se repetem mais vezes. Este é um método útil para extrair dados em páginas com muitos registros.

Wang and Lochovsky (2002) propõem o *Data-rich Section Extraction* (DSE), que é um algoritmo de extração de regiões ricas de dados em páginas de um mesmo site que utilizam um template comum. Páginas de um mesmo site são transformadas na árvore de tags DOM para realização das comparações entre árvores, dessa forma, são identificadas as que possuem a mesma estrutura. Subárvores que compartilham as mesmas informações em árvores de tags diferentes são removidas (região que se repete entre as páginas do site). O que resta é a região rica em dados; com isto, pretende-se eliminar menus, propagandas e outros elementos comuns a todas as páginas do site. É um método que não extrai apenas registros, mas, assim como o RoadRunner, necessita realizar comparações entre páginas que dividem

o mesmo layout/template.

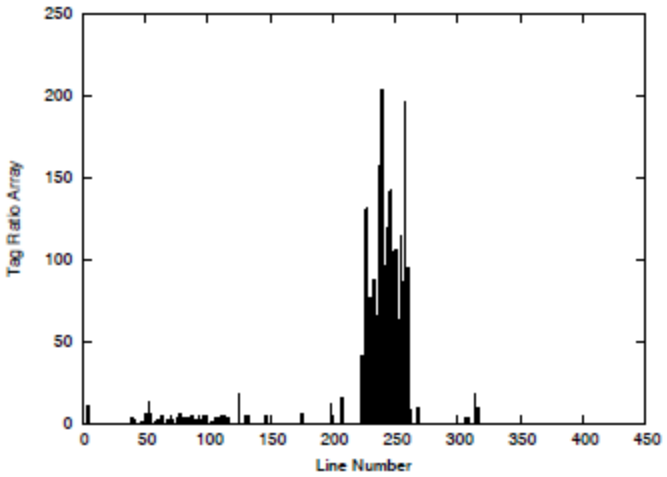
O trabalho apresentado em VIPS (CAI et al., 2003) propõe a divisão da página em blocos, percorrendo a árvore DOM de cima para baixo, construindo uma nova árvore de blocos com base nas informações visuais da estrutura da página WEB. Assim, a página WEB é separada em blocos determinados por informações visuais da árvore DOM original, como tamanho da fonte, quebra de linhas, cor de fundo. Método similar, mas mais específico para artigos e necessitando de treino da base de dados para dividir diferentes partes de uma página HTML com informações visuais é realizado por Burget and Rudolfova (2009), que extraem partes do texto de base de artigos que têm seus blocos classificados de acordo com suas informações visuais (fonte, característica de cor, de borda e outras) e posição espacial na página. Este método permite detectar algumas regiões de artigos como subtítulo, parágrafo, data e autor. Intitulado WEB Page Element Classification Based on Visual Features, o abreviaremos por CVF.

O método proposto por Insa et al. (2013), intitulado *Words/Leafs Ratio* – WLR -, percorre toda a árvore DOM em busca de subárvores que possuam maior concentração de texto. Este método não necessita de uma página modelo, nem conhecer outras páginas do mesmo site. Para isto, exclui os nós *title*, *script*, *comment* e os nós folhas que não possuem conteúdo-texto, além de unir nós filhos de um mesmo nó, os quais geralmente representam uma única porção de texto, sendo representados por nós irmãos, que possuem as tags *p, a, u, b, i, em, span, sub, sup, strong* e *div*. Uma função de Relevância que leva em consideração os nós com maior densidade de texto nele e em seus filhos é responsável por escolher a subárvore de maior relevância.

A abordagem *Content Extraction via Tag Ratio* (CETR) (WENINGER et al., 2010) percorre todo o código HTML da página WEB, linha a linha, construindo um vetor cujo índice é a linha e, o conteúdo, a razão de quantidade de caracteres por tag de cada linha. Não havendo tag, o conteúdo é a quantidade de caracteres. Após este procedimento, constrói-se um gráfico/histograma no qual é possível visualizar as áreas das páginas com maior concentração de conteúdo textual, conforme Figura 1 – Histograma CETR (WENINGER et al.,

2010). Depois disso, utilizam-se algoritmos de *clustering* para agrupar o gráfico em grupos de dados; em seguida, escolhe-se o grupo com maior quantidade de conteúdo. Esta abordagem é muito útil e apresenta bons resultados na extração automática de ruídos como menus, propagandas e outras áreas não relacionadas ao conteúdo textual, além de estar disponível na WEB, toda a implementação do método em código-fonte java.

Figura 1 – Histograma CETR



Fonte: Weninger et al., 2010.

Apesar da divisão das técnicas de extração nos grupos acima, com base em suas características principais, uma técnica pode utilizar aspectos relacionados a outras. O TPS, por exemplo, verifica os estilos das tags para diferenciação de trilhas de tags com o mesmo "tag-path"; o CETR parte da premissa de que a maioria das páginas contém um pequeno banner no topo, uma lista de links à esquerda ou à direita e o conteúdo principal na região do meio. A técnica de extração proposta no presente trabalho, Query-based Blocks Mining (QBM), se aproxima dos trabalhos do grupo (ii) por definir blocos HTML, aninhados ou não, e a partir deles excluir os blocos considerados irrelevantes.

2.2.2 Análise Comparativa

Quando comparadas aos trabalhos citados, podem-se destacar as características apontadas na tabela 2. Sendo:

- "Depende de HTML", são métodos que se baseiam nas tags específicas HTML para segmentação da página em blocos ou extração da região principal do documento HTML;
- "Direcionado à extração de registros", são métodos cujo processo de segmentação da página ou extração da região principal do documento HTML é voltado para identificação de registros de dados, sejam eles simples ou aninhados. Registros de dados são regiões, da página html, de dados semiestruturados que podem ser considerados linhas de informações tabulares;
- "Página Modelo", são métodos cujo processo de segmentação ou extração da região principal do documento HTML necessita de uma ou algumas páginas modelos que seguem layout ou estrutura html semelhante para realização de comparações entre a página que se deseja extrair dados e o modelo comparado e posterior segmentação ou extração de dados;
- "Extração Direcionada à Consulta", caso que os termos da consulta do usuário influenciam no processo de segmentação ou extração da região principal do documento HTML;

Tabela 2: Algoritmos de Extração de dados de páginas Web

Trabalho	***Dependente de HTML***	***Direcionado à extração de registros***	***Página Modelo***	***Extração Direcionada à Consulta***
VIPS	Sim	Não	Não	Não
CETR	Não	Não	Não	Não
WLR	Não	Não	Não	Não
TPS	Não	Sim	Não	Não
MDR	Não	Sim	Não	Não
Road Runner	Não	Não	Sim	Não
DSE	Não	Não	Sim	Não
CVF	Sim	Não	Não	Não
QBM	Sim	Não	Não	Sim

Fonte: Autor

A proposta apresentada possui a vantagem de se voltar especificamente à busca efetuada, focando apenas nas regiões que possuem conteúdo relacionado aos termos da consulta, excluindo as demais áreas, além de não necessitar de página modelo para identificação de regiões principais e não ser específico a para extração de registros semiestruturados.

Como o trabalho proposto tem o intuito de apresentar um método de ranking de páginas HTML que satisfazem à consulta efetuada, executado sob uma base local com páginas apresentando formatos (layout) variados, independente de serem páginas de e-commerce, fóruns, bibliotecas (wikipedia por exemplo), podendo o conteúdo principal ser apresentado de forma tabular ou não, faz com que neste contexto específico a distribuição de melhor configuração de valores das colunas da Tabela 2 para o método de filtro ou segmentação capaz de identificar a região de interesse para usuário a ser aplicado seja {sim, não, não, sim}.

2.3 RANKING DE DOCUMENTOS HTML

Algoritmos de ranking de documentos HTML podem ser construídos levando-se em conta vários aspectos. Selvan et al. (2012) propõem três categorias para algoritmos de ranking:

(i) a primeira, **baseada na análise dos links**, trata de trabalhos focados na análise de links de um conjunto de documentos para definição do ranking; são exemplos PageRanking (PAGE et al., 1999) e HITS (KLEINBERG, 1999);

(ii) a segunda, **baseada na busca personalizada**, considera a consulta do usuário ou aspectos de feedback fornecidos por ele; entre eles, ClickThroughData (JOACHIMS, 2002), Page Content Ranking (POKORNY and SMIZANSKY, 2005) e o clássico modelo vetorial utilizado pelo Lucene (HATCHER et al., 2004) - nesta categoria, Duhan et al. (2009) usam o termo WUM

(WEB Usage Mining) para identificar estes trabalhos;

(iii) a terceira, **baseada na segmentação da página**, consta de algoritmos que dividem a página em blocos, da qual se podem citar: FixedPs (CALLAN, 1994), Incorporating Window-Based Passage-Level Evidence in Document Retrieval (XI, 2001), Block-Based WEB Search (CAI et al., 2004) e Computing Block Importance for Searching on WEB Sites (FERNANDES et al., 2007).

O clássico algoritmo de PageRanking (PAGE et al., 1999) utiliza uma técnica de ranking baseada no relacionamento entre diversas páginas WEB. Um crawler realiza o download de centenas de milhares de páginas WEB. Através dos links destas páginas, são baixadas mais páginas; quanto mais uma página for referenciada por outras, maior será o ranking desta página. Este algoritmo é completamente automatizado e muito útil para definição do ranking inicial de um grande conjunto de páginas WEB, sem necessidade de interação com o usuário.

Similarmente, o HITS (KLEINBERG, 1999), desenvolvido para quantificar o valor de authority e do hub de uma página, também calcula o ranking com base na análise de links. Uma página tem o valor de authority alto quando esta é apontada por muitas outras páginas (hubs) e o valor de hub é alto quando aponta para várias outras páginas (authorities). Uma máquina de busca retorna às páginas WEB que mais satisfazem uma determinada consulta, o que é chamado conjunto raíz R; neste subconjunto R de páginas, aplica-se o algoritmo HITS (DEVI et al., 2014).

Em Joachims (2002), a proposta é o uso de uma função de aprendizado de máquina chamada clickthrough data, que utiliza especificamente as informações dos links acessados (clicados) pelo usuário para tornar estas páginas visitadas mais prioritárias que as demais. O método proposto em Pokorny and Smizansky (2005), intitulado Page Content Ranking (PCR), avalia a proximidade da página WEB com os termos da pesquisa efetuada, baseado em características como frequência dos termos, número de páginas que contêm o termo e a ocorrência de sinônimos, realizando a comparação do PCR com o

PageRank. O PCR utiliza uma rede neural para detectar a importância de uma página para uma dada busca, o que requer treinamento da rede e, consequentemente, interação do usuário.

Nesta categoria baseada na busca personalizada, também pode-se citar o algoritmo de ranking do Lucene (HATCHER et al., 2004), que usa o Modelo Vetorial (VSM) ou o Modelo Booleano, para determinar a relevância de um dado documento em relação a uma consulta específica de um usuário. Este modelo é amplamente utilizado em ferramentas de máquinas de busca, como o Lucene da Apache Software Foundation, que possui implementação disponível em código java e não utiliza aprendizado de máquina.

Callan (1994) divide a página em janelas de tamanhos entre 50 e 300 palavras, unindo parágrafos pequenos e subdividindo, quando maiores. Neste estudo, ele conclui que realizar o ranking em parágrafos (janelas) de tamanho fixo é melhor que realizar o ranking de passagens obtidas pelas sentenças limite dos parágrafos e que, apesar de eficiente o ranking em janelas de 50 a 300 palavras, o tamanho ideal irá depender de cada base de documentos.

De maneira similar, Xi (2001) combina o ranking baseado em janelas de tamanho fixo com o ranking do documento inteiro para obter um melhor ranking final. A ideia é que um parágrafo ou seção importante para determinada consulta possa não ser retornado se o documento, como um todo, for irrelevante para determinada consulta. A ideia é que realizar a similaridade da consulta com pequenas partes de documentos muitos longos pode melhorar o ranking. Estes métodos são utilizados em documentos textuais, avaliando *df* e *idf* como critérios de ranking. Em (CAI et al., 2004), a página é dividida em blocos, com base em janelas de tamanho fixo, no algoritmo VIPS, e em uma combinação dos dois; após isto é calculado o ranking destas páginas WEB analisando as janelas criadas através destes três diferentes métodos. O melhor resultado é obtido com a combinação do algoritmo VIPS, que gera blocos de páginas HTML, tornando este de tamanho fixo, retornando como escore o BM25 do melhor bloco avaliado em cada página.

Fernandes et al. (2007) utilizam o método VIPS para dividir em

blocos páginas com estruturas similares (podendo os blocos ser divididos por outro método, ou mesmo manualmente), criando o conceito de *page class (classe de páginas)* que são páginas WEB com estrutura HTML equivalentes e *block class (classe de blocos),* ou conjunto de blocos similares de cada *page class.* Este algoritmo assume que quanto mais blocos, em uma mesma página, um termo possui, mais relevante ele é para o assunto principal da página, utilizando uma adaptação da medida *idf*, para, no lugar de considerar a frequência inversa de termos nos documentos, fazer a medida por blocos. Assim, quanto mais blocos de uma classe de blocos possuem conteúdo igual, ou muito similar, o *icf* (inverse class frequence) dos termos dessa classe de blocos será baixo e a classe de blocos que seu conteúdo não se repete entre as páginas, terá seu valor médio de *icf* elevado. Em uma mesma página, blocos de classes diferentes, mas que possuem termos em comum com outros blocos dentro da mesma página, tenderão a ter seu assunto relacionado ao conteúdo principal da página, gerando o conceito de *blockSpread,* que será maior quanto mais termos em comum este bloco tiver com outros da mesma página, fazendo com que o *classSpread,* média dos blockspread dos blocos da classe, seja alto. Assim o ranking da página é avaliado em função do *aicf* (média do icf) e do *classSpread* dos blocos que possuem os termos da consulta, produzindo melhores resultados que o ranking do BM25 aplicado aos mesmos documentos.

Estes métodos não consideram aspectos específicos de documentos HTML na etapa de ranking; não diferenciando se os termos estão em destaque, no título ou em descrição de links.

O método proposto neste trabalho realiza um processo de segmentação; ao mesmo tempo, utiliza a consulta do usuário para a melhoria do ranking, possuindo assim fortes aspectos de (ii) ou (iii).

2.2.3 Análise Comparativa

A tabela 3 contém as características dos algoritmos de ranking. Sendo:

- “Aspectos HTML no ranking”, analisa se o método de ranking

necessita do uso de tags específicas HTML para definição da relevância. Quando o método avalia informações como se o termo está em destaque, no título, como descrição de links, em itálico ou quantidade de links de uma página web, por exemplo, necessitará reconhecer determinadas tags HTML, mesmo que utilize alguma biblioteca que abstraia o reconhecimento destas tags;

- "Personalização", analisa se o método de ranking necessita reconhecer o usuário que está realizando a consulta para que através de suas preferências de pesquisa anteriores possa alterar o peso das páginas web de maneira específica baseado nas características pessoais de navegabilidade coletadas deste usuário, retornando um ranking personalizado;
- "Proximidade documento e consulta", analisa se o método de ranking depende dos termos da consulta do usuário para determinar a relevância do documento analisado;
- "Aprendizado de Máquina" analisa se o método de ranking utiliza aprendizado de máquina o que requer uma base grande para treino e maior complexidade computacional para determinação do melhor ranking.

Tabela 3 - Algoritmos de Ranking na WEB

Trabalho	**Aspectos HTML no Ranking**	**Personalização**	**Proximidade documento e consulta**	**Aprendizado de Máquina**
Page Ranking	Sim	Não	Não	Não
PCR	Não	Não	Sim	Sim
VSM	Não	Não	Sim	Não
Callan (1994)	Não	Não	Sim	Não
Xi (2001)	Não	Não	Sim	Não
Click Through Data	Sim	Sim	Não	Sim

(cont.)

Continuação)

Block-Based Web Search	Não	Não	Sim	Não
Block Importante on Web Site	Não	Não	Sim	Não
FPR	Sim	Não	Sim	Não

Fonte:Autor

O FPR é um algoritmo de ranking de segmentação da página com mineração de conteúdo (WCM), ou seja, extrai o conteúdo da página HTML para determinar sua relevância para determinada consulta, especificamente utilizando-se da mineração de texto (Web Text Mining), não sendo objeto do algoritmo a mineração de imagens e vídeos.

Pode-se destacar ser este mais específico que os algoritmos PageRanking e HITS, permitindo um ranking próprio para uma consulta específica. O FPR é voltado a documentos HTML; avalia aspectos específicos de tags HTML que Pokorny and Smizansky (2005), Callan (1994), Xi (2001), Cai et al. (2004) e Fernandes et al. (2007) não consideram, como termos em destaque, ocorrência no título e palavras-chaves em links. Além disso, tem a vantagem, sobre os métodos de ranking - PageRanking, HITS, PCR, Vector Space Model, Clickthrough Data, de realizar a extração de ruídos antes da etapa de ranking.

Com relação ao uso de inteligência artificial, esta pode ser utilizada para determinação dos valores fixos dos pesos dos critérios de ranking, após isto, o processo de ranking deve seguir sem o uso de inteligência artificial, o que se tornaria muito custoso para uma base de dados locais de páginas HTML que não é estática (pode crescer) além de se ter um número imenso de diferentes consultas efetuadas e a necessidade de se obter o ranking em tempo hábil para o usuário.

O uso de personalização para cada usuário no método de ranking exige que as páginas Web mais acessadas pelo usuário sejam

reconhecidas e as preferências de cada usuário cadastrado sejam armazenadas pelo método de ranking, o que não faz parte do escopo deste trabalho, podendo isto ser feito com agregação de ranking para definição de um ranking final, o método proposto FPR além de ser voltado para páginas HTML é voltado para consulta efetuada, mas não realiza o reconhecimento do usuário, podendo este entrar como ranking inicial caso seja preferível o ranking de páginas HTML personalizado para cada usuário.

Sendo assim, no contexto de um processo de ranking específico de páginas HTML voltado à consulta efetuada em uma base local, utilizada por usuários não previamente cadastrados e que deve retornar o resultado em tempo hábil, temos a melhor configuração de valores das colunas da tabela 3 como sendo {*sim, não, sim, não*}.

3 FILTERED-PAGE RANK

O método de ranking proposto, chamado FPR (**Filtered-Page Ranking**), possui duas etapas principais:

(i) *Segmentação da página Web*; e

(ii) (ii) *ranking dos documentos*.

O objetivo do processamento executado na etapa de segmentação é identificar os blocos relevantes para consulta que serão utilizados na etapa de ranking de documentos HTML e com isto eliminar do processo de ranking nodos que não sejam relevantes para a busca (menus, imagens, informes, entre outros, contendo texto não relacionado às palavras-chave da consulta). Durante a etapa de ranking, o foco é a ordenação dos documentos relevantes para uma dada consulta, usando certos critérios para quantificar quão próximo o conteúdo de uma página é dos termos da busca.

3.1 VISÃO GERAL

O processo geral de coleta das páginas Web e ranqueamento é bastante similar ao funcionamento de um sistema de recuperação de informação tradicional. O procedimento de coleta das páginas pode ser realizado por qualquer crawler. As páginas são armazenadas de tal forma que dados e metadados são mantidos em um repositório para posterior consulta. Nesta etapa, são armazenados metadados com as informações da árvore DOM original, mapeando os termos com os nodos relacionados e demais propriedades do nodo, tais como a tag HTML, quantidade de vezes que o termo aparece no nodo, entre outros. Estes dados são parâmetros para o algoritmo de ranking. Apenas um subconjunto de nodos da árvore DOM original, chamada de DOM filtrada, é analisado, não sendo foco deste trabalho a forma de armazenagem e a recuperação destes metadados.

De forma geral, a partir da(s) palavra(s) submetida(s) a consulta, o processador analisa os metadados presentes na árvore DOM filtrada, utilizando-os para quantificação da relevância da página filtrada, calculando quão próximo o conteúdo é da consulta. Finalmente,

os resultados são exibidos de forma decrescente de relevância. Na fase de consulta, o usuário pode especificar termos que obrigatoriamente devem aparecer nas páginas. Se a página não tiver alguma palavra obrigatória, ela não é retornada.

3.2 QUERY-BASED BLOCKS MINING

O QBM é a etapa no qual o documento HTML é segmentado em blocos. Os blocos delimitam regiões que terão tratamento específico para definir o que é conteúdo relevante ou não. O objetivo desta fase é extrair uma árvore DOM filtrada que só possua blocos diretamente relacionados à busca do usuário, descartando conteúdo irrelevante.

O método QBM proposto foi criado para identificar em páginas HTML, que utilizam o HTML 5.0 ou inferior, blocos relacionados às palavras-chaves da consulta efetuada, permitindo que métodos de indexação ou ranking foquem seu processamento somente em partes da página Web consideradas relevantes para determinadas palavras-chaves.

O objetivo do método de extração proposto é melhorar a precisão do FPR na etapa de ranking de páginas HTML. Etapa esta que avalia a proximidade de um documento com uma consulta efetuada, através da análise de critérios utilizados em estudos de recuperação da informação apresentados na seção 3.3.

3.2.1 Categorização dos blocos

Para que esta tarefa seja executada, primeiramente o documento HTML coletado é tratado como uma árvore DOM, cujos nodos são categorizados em três grupos: (i) *blocos-segmento*; (ii) *blocos-descarte* e (iii) *blocos-destaque*, cujas categorias são utilizadas para delimitar conteúdo útil e inútil ou em fase de ranking.

Definição 1. *(Árvore DOM categorizada): Seja* $N = \{n_1, \ldots, n_i\}$ *um conjunto de nodos e* $A = \{a_1, \ldots, a_{i-1}\}$ *o conjunto de arestas que liga os nodos de N. Uma árvore DOM categorizada é definida como um par ADOM = (N, A), onde N representa um conjunto de nodos cujo* n_j *pode*

representar Blocos-Segmento, Blocos-Destaque ou Blocos-Descarte.

Uma árvore DOM categorizada, que representa o documento HTML original, conforme coletado da WEB, possui tanto nodos importantes para o processo de ranking quanto nodos a serem eliminados. Estes nodos podem representar blocos-segmento, blocos-destaque ou blocos-descarte, que são tratados conforme definição abaixo.

Definição 2 *(Bloco-Segmento). Seja DC uma árvore DOM categorizada e* n_j *um nodo qualquer de DC. Um bloco Bsg =* n_j *é uma subárvore de DC, chamado bloco-segmento, tal que* n_j *seja qualquer região contínua de texto, Bsg* $\subset DC$.

Blocos-segmento são subárvores da árvore DOM categorizada, capazes de delimitar regiões; ou seja, blocos-segmento são elementos que o FPR considera capazes de delimitar contexto (agrupando elementos HTML, ou conjunto de palavras que precede ou segue as palavras-chave da consulta), podendo um bloco-segmento estar contido em outros. Estas regiões podem indicar blocos nos quais o termo consultado está inserido, ou conter blocos que possuem ruído e devem ser eliminados. Geralmente, delimitam regiões contínuas de texto ou inseridas dentro de um contexto que as agrupe, caso da tag form ou div, que define um conjunto de dados de um formulário ou o estilo e formatação, respectivamente. Blocos-segmento podem ser representados, por exemplo, pelas tags *{html, body, form, div, table, tr, iframe, article, section, ul, li, title, meta}*.

Definição 3 *(Bloco-Destaque). Seja* n_j *um nodo qualquer, que pode conter um elemento HTML de formatação de caractere. Um bloco Bdq =* n_j *é uma subárvore de DC, chamado bloco-destaque, tal que Bdq* $\subset$ *Bsg e Bdq sejam representados por um nodo que contenha um elemento de formatação de caractere.*

Blocos-destaque são blocos especiais que contêm elementos HTML de formatação de caractere, ou seja, formatam ou destacam determinados pedaços de texto, podendo, por exemplo, sublinhar, marcar em negrito ou itálico e alterar o tamanho da fonte. Sempre está contido em um *bloco-segmento* e não delimita regiões que o FPR considera segmento de texto; apenas destaca partes de regiões contínuas

de texto. O *bloco-destaque* não é considerado na etapa de extração, só sendo preservado se seu bloco-segmento ascendente mais próximo também for preservado. Os blocos-destaque são importantes na etapa de ranking para determinar o quanto o conteúdo texto da página HTML é próximo dos termos da consulta. Podem ser representados, por exemplo, pelas tags *{strong, b, i, u, span, a, h1, h2, h3, h4, h5, h6}*.

Definição 4 *(Bloco-Descarte). Seja* n_j *um nodo qualquer, que pode conter um elemento vazio, invisível ou oculto. Um bloco Bdc =* n_j *é uma subárvore de DC, chamado bloco-descarte, tal que Bdc* $\subset$ *DC, e Bdc é um elemento vazio (não contém texto nem subárvores) ou invisível ou oculto.*

Blocos-descarte são blocos de eliminação automática, pois não têm conteúdo texto visível (são vazios, ocultos ou invisíveis). Toda subárvore é excluída automaticamente, por ser bloco-descarte, quando: (i) o nodo representa um elemento vazio, ou seja, não possui texto em si; (ii) o nodo representa um elemento que tem semântica de oculto, ou invisível, não aparecendo na apresentação da página HTML, contendo, por exemplo, atributos iguais a "*style*" = "*display: none*";"*visibility*" = "*hidden*" e "*visibility*" = "*collapse*".

3.2.2 Criação da árvore filtrada

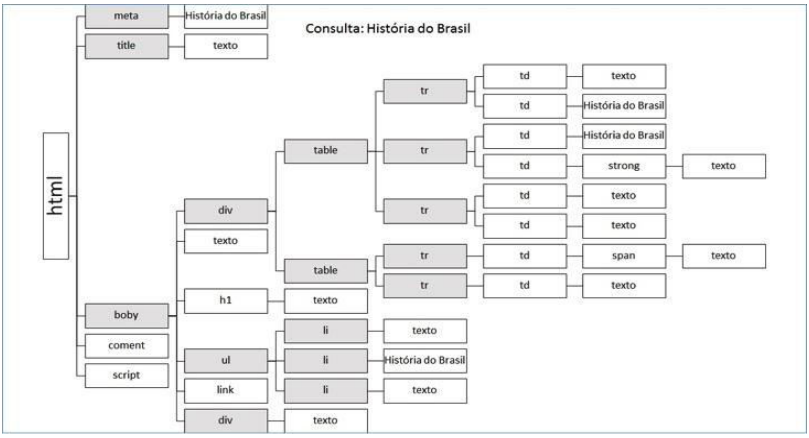
Em uma árvore DOM categorizada, o nodo-pai é o bloco-segmento principal, que pode ser formado por diversos outros blocos-segmento (conforme se pode observar na figura 2). Os blocos-segmento que possuem alguma palavra-chave da busca do usuário formam a árvore DOM filtrada.

Definição 5 *(Árvore DOM filtrada). Seja DC uma árvore DOM categorizada,* $C = \{t_1, t_2,...,t_n\}$ *o conjunto de palavras-chaves da consulta do usuário,* $BDC = \{bdc_1, bdc_2,...bdc_n\}$ *o conjunto de blocos-descarte da árvore original e* $BSG' = \{bsg_1, bsg_2,...bsg_n\}$ *o conjunto de blocos-segmentos, tal que* $BSG' \not\subset t_i$. *Uma* ***árvore DOM filtrada*** *(Af) é uma árvore, tal que Af = DC – BDC – BSG'.*

Uma árvore DOM filtrada é composta por blocos-segmentos que contenham alguma palavra-chave da busca do usuário, sem os blocos-descarte que já foram automaticamente eliminados. Para a geração da árvore filtrada, os blocos-segmento que não possuem qualquer uma das palavras-chave da consulta são descartados. Em blocos-segmentos aninhados, os blocos filhos que não possuem nenhuma palavra-chave são excluídos, preservando o bloco-segmento pai se este ou pelo menos um bloco-segmento filho tiver pelo menos uma palavra-chave da busca. Em blocos aninhados, a palavra pertence a determinado nodo n_j e a todos os nodos ascendentes de n_j, mas não pertence aos nodos descendentes, ou irmãos de n_j na árvore DOM.

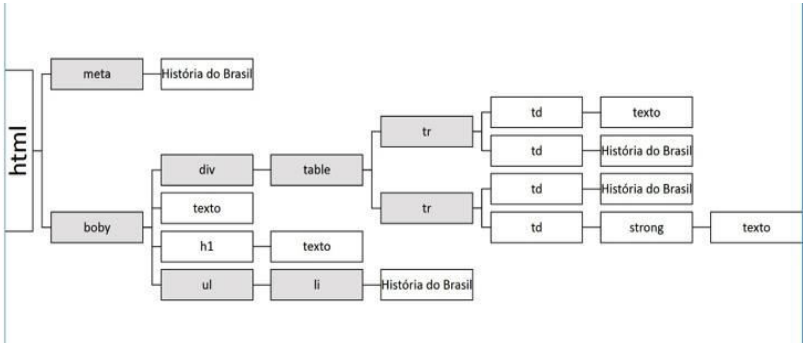
A figura 2 mostra um exemplo de uma árvore DOM categorizada, com blocos-segmento, blocos-destaque e blocos-descarte. A figura 3, uma árvore filtrada sem os nodos excluídos, por serem nodos-descarte (por exemplo, o nodo script) ou por serem blocos-segmento e não conterem algum termo da busca. Os blocos-segmento são os retângulos com fundo preenchido.

Figura 2 - Árvore DOM antes QBM



Fonte: Autor.

Figura 3 - Árvore DOM após QBM



Fonte:Autor

3.3 RELEVÂNCIA DE UMA PÁGINA HTML

Para a realização do processo de ranqueamento, o algoritmo utiliza alguns parâmetros cujos valores são obtidos na árvore filtrada ou nos metadados armazenados. Alguns critérios são baseados em um estudo realizado por pesquisadores da área de Ciência da Informação (BORGES, 2009), que afirmam que para a compreensão do conteúdo do documento a leitura integral é ideal, embora seja impraticável.

Neste estudo, que levou em conta diversos trabalhos da comunidade de recuperação de informação, definem-se os trechos de documentos e critérios que deveriam ser considerados mais importantes para a indexação de documentos em meio digital, dos quais o FPR utiliza: (i) **título**, (ii) **introdução e primeiras frases de capítulos/parágrafos**, (iii) **tabelas e listas**, (iv) **palavras em destaque**, (v) **frequência dos termos,** (vi) **palavras vazias de significado** e (vii) **frases-termo**. Além dos critérios baseados neste estudo, o FPR utiliza ainda (viii) **quantidade de links com todas as palavras-chave da busca em sua descrição** e (ix) **quantidade de palavras** para definir quão próximo uma página HTML é de determinada consulta.

A adaptação feita no FPR para definição dos critérios de relevância deve levar em conta, também, os elementos HTML. Cada critério é tratado da seguinte forma:

(i) **título**: é uma fonte importante de extração das palavras-chave de um documento; em uma página HTML; consideram-se as tags *title* e *meta* que seguem o mesmo raciocínio e são *blocos-segmentos*;

(ii) **introdução e primeiras frases de capítulos e parágrafos**: este fato é baseado na observação que as primeiras frases de capítulos e parágrafos, em geral, possuem os termos mais relevantes para um processo de indexação, o algoritmo de extração QBM parte de premissa similar e infere que o conteúdo relevante para a definição do *ranking* está próximo dos termos da consulta, contido no mesmo bloco-segmento de algum termo da consulta;

(iii) **tabelas e listas**: são consideradas blocos-segmento, e todo seu conteúdo é levado em conta, podendo ser representados, por exemplo, pelos elementos *table, ul/ol, tr e li*;

(iv) **palavras em destaque:** são palavras enfatizadas no texto, e podem ser sublinhadas, em negrito ou destacadas, com tamanho ou fontes diferentes, e são consideradas para pontuação diferenciada, aumentando a relevância de um documento; estão contidas em *blocos-destaque*;

(v) **frequência dos termos**: quanto mais vezes um termo da busca efetuada aparecer no documento, maior será a relevância do documento;

(vi) **palavras vazias de significado**: palavras que não influem na relevância do documento e prejudicam o método de *ranking, filtro* ou indexação caso sejam consideradas no processo, pois são vazias de significado e representam artigos, preposições ou conjunções que não são consideradas palavras-chave/termos da busca efetuada; são as chamadas *stop-words* (BAEZA, 1999).

(vii) **frases-termo:** representam palavras mais específicas, inseridas dentro de um mesmo contexto, não devendo ser tratadas de forma isolada, os termos da consulta são tratados como frases-termo.

Ademais, quanto menos texto, menos relevante tende a ser a página Web que satisfaz determinada consulta, assim a quantidade de palavras é considerada como critério e em páginas HTML links/âncoras representam um importante elemento da linguagem HTML e permite o direcionamento entre páginas ou para uma parte específica na mesma página HTML com isto a ocorrência de palavras-chaves em links deve ser tratada de forma diferenciada. O método proposto FPR armazena as informações dos links das páginas HTML em metadados como propriedades dos blocos-segmento. Optou-se por não tratar link como bloco-destaque no algoritmo QBM, pelo fato de só considerar este para atribuir uma pontuação maior para página HTML caso este possua todas as palavras-chaves e atribuir um peso maior para esta ocorrência. Todos os blocos-destaque possuem função de realçar, em geral, pequenas porções de texto e link função de navegabilidade. Além disto, blocos-destaque possuem o mesmo tratamento de considerar cada palavra-chave isoladamente e todos terem o mesmo peso no algoritmo FPR. Por estes motivos as informações dos links são armazenadas em metadados e avaliadas no processo de ranking e não consideradas blocos-destaque. Em trabalhos futuros sugerimos um estudo específico na definição de pesos ótimos, podendo ser atribuído um peso para cada tipo de bloco-destaque e links serem tratados como tal.

Segundo Borges (2009): “Em geral, não é necessário armazenar as palavras de forma composta, pois este processo de unificação das palavras exige tempo”. Salton (1983) e Croft (1982) recomendam que ela não seja utilizada, pois não aumenta de forma considerável a eficiência do sistema. O que pode ser feito é o armazenamento da informação sobre as distâncias entre as palavras de um mesmo documento e deixar que a técnica de consulta avalie se as palavras são ou não adjacentes. O FPR penaliza páginas cujos termos da consulta estão separados por uma distância maior que um valor pré-determinado *d*, definido na seção 4 – Avaliação Experimental, por inferir que não fazem parte de um mesmo contexto. Este dado é obtido pela observação segundo a qual, em páginas relevantes, as palavras-chaves da consulta se encontravam separadas, em sua maioria, por uma distância máxima *d* e

que as páginas irrelevantes com muito conteúdo texto possuem, na maioria dos casos, distâncias mínimas entre as palavras-chave maior que *d*. Esta característica aumenta a precisão do FPR, passando a considerar corretamente irrelevantes algumas páginas que antes eram tratadas como relevantes, não se observando o oposto, que é o de passar a considerar irrelevantes páginas relevantes. A quantidade de links com todas as palavras-chave em sua descrição aumenta o ranking de uma página, pois links são elementos de destaque de páginas HTML e desempenham papel importante na concepção das páginas WEB, direcionando o usuário para a parte da página, ou a outra página, que trata do assunto da descrição do link, sendo considerada uma fonte valiosa para se encontrar a informação desejada.

Os experimentos mostraram que, de forma geral, as páginas WEB são relevantes por possuírem uma quantidade de texto significativa sobre o assunto consultado ou uma quantidade mínima de links com as palavras-chave da consulta em sua descrição; com isto, páginas WEB que não possuem uma quantidade mínima de palavras não vazias de significado ou que não tenham uma quantidade mínima de links com todas as palavras-chaves obrigatórias em suas descrição serão penalizadas no seu ranking final. Esta característica aumentou a precisão do FPR e será chamada de *função de conteúdo mínimo,* ou $f_{min,}$ para abreviar.

3.4 FUNÇÃO DE RANKING E O ALGORITMO FPR

A função de ranking é definida levando em conta os critérios de relevância descritos na seção anterior, considerando a importância de certas partes do documento (título e destaques, por exemplo), bem como o número de ocorrência dos termos da consulta em certas partes do documento.

Antes de introduzir a função de ranking, é importante definir a correlação entre as palavras-chave da consulta. A pesquisa em questão considera que os termos não são mutuamente independentes, ou seja, casos em que t_i não possui relação nenhuma com $t_{i+1,}$ o que não ocorre na maioria das consultas; por exemplo, as palavras "*recuperação*" e

"*informação*" tendem a aparecer próximas em documentos sobre sistemas de recuperação da informação (BAEZA-YATES 199, RIBEIRONETO 1999). Nestes documentos, o aparecimento de uma palavra-chave atrai o surgimento de outra, pois são termos que possuem correlação. Nesta pesquisa, refletimos sobre esta relação, que é medida pela distância entre as palavras-chave em cada documento de acordo com o seguinte:

Definição 6 (*Função correlação*). *Seja Seja C = {t_1, ..., t_n} o conjunto de palavras-chaves da consulta, DC uma árvore categorizada DOM (filtrada ou não), a correlação de C em DC é calculada pela seguinte função:*

$$D(C, DC) = \begin{cases} 1, & d(ti, tj) < d \\ \alpha, & d(ti, tj) \geq d \end{cases}$$

Nesta definição, *d* é um parâmetro que indica a distância máxima entre as palavras-chave da consulta na página WEB e α um valor menor que 1, usado para penalizar quando a distância entre os termos da consulta é maior que *d;* ou seja, se há termos de C na página WEB distante mais que *d* caracteres de qualquer outro termo de C, então o ranking desta página é penalizado.

Definição 7 (*Função F_{min}*). *Seja Seja C = {t_1, ..., t_n} o conjunto de palavras-chaves da consulta, DC uma árvore DOM categorizada (filtrada ou não) e f_{tt}(DC) uma função que retorna o total de termos em DC e L = {l_1, ..., l_k} o conjunto de k links com todos os termos de C em sua descrição, a função de conteúdo mínimo é calculada pela seguinte função:*

$$Fmin(C, DC) = \begin{cases} 1, & ftt(DC) > x \; ou \; k > y \\ \alpha, & caso \; contrário \end{cases}$$

Nesta definição, *x* é um parâmetro que indica a quantidade mínima de termos da página WEB; *y,* a quantidade mínima de links com

todas as palavras-chave na descrição e α um valor menor que 1, usado para penalizar quando a página WEB possui menos de *x* palavras não vazias de significado, ou menos de *y* links com as palavras-chave na descrição.

Definição 8 (*Função de ranking*). *Seja* $C = \{t_1, ..., t_n\}$ *o conjunto de palavras-chaves da consulta, DC uma árvore DOM categorizada e* $L = \{l_1, ..., l_k\}$ *o conjunto de k links com todos os termos de C em sua descrição. O grau de relevância R(C, DC) de DC para a consulta C é calculada pela seguinte função:*

$$R(C,DC) = D(C,DC) * Fmin(C,DC) * \left\{ \left(W_1 * \sum_{i=1}^{n} f_o(t_i) \right) + \left(W_2 * \sum_{i=1}^{n} f_{hb}(t_i) \right) + (W_3 * k) + \left(W_4 * \sum_{i=1}^{n} f_{tm}(t_i) \right) + (W_5 * f_t(A)) \right\}$$

Onde:
D(C; DC) é a função de correlação descrita na definição 6;
Fmin(C; DC) é a função de conteúdo mínimo descrita na definição 7;
$f_o(t_i)$ é o número de ocorrência do termo t_i na árvore *DC;*
$f_{hb}(t_i)$ é o número de ocorrência do termo t_i em blocos destaque de *DC*;
k é o total de links com todas as palavras-chaves na descrição;
$f_{tm}(t_i)$ é o número de ocorrência do termo t_i nas tags *title* ou *meta* da *DC*;
f_t é o total de palavras em *DC* e
Wi: indica o peso atribuído a cada critério.

A função *R(C,DC)* tem o intuito de calcular a proximidade de uma página HTML *DC* com a consulta *C,* utilizando os critérios de relevância listados na seção 3.3. A cada critério é atribuído um peso, representado por *Wi*, conforme podemos constatar na definição 8.

O algoritmo que implementa o método de ranking FPR é apresentado na figura 4 em pseudo-code, com os passos de (i)

segmentação/remoção e (ii) *ranking*. Nas linhas 2 a 4, valores de relevância são atribuídos a cada documento HTML da coleção que satisfaz a consulta (linha 1). A função de extração de conteúdo irrelevante, chamada *segmentRemoval* (linha 23), segmenta a página em diferentes categorias de blocos e remove os blocos-descarte (definição 4) e os blocos-segmentos (definição 2) que não possuem termos da consulta efetuada (linhas 26 a 33); então, o resultado da relevância de cada página filtrada é calculado com base em critérios predeterminados (linhas 11 a 21) e, finalmente, retornado em ordem decrescente de relevância (linha 5).

Figura 4 - Algoritmo FPR

Algorithm 1 Ranking algorithm (FPR)

```
1: INPUT C
2: for all dom ∈ base() do
3:     rank[name, value] = [Name(dom), Relevance(dom, C)]
4: end for
5: OUTPUT OrderDesc(rank)
6: procedure RELEVANCE(dom, C, filter)
7:     Af ← dom
8:     if (filter) then
9:         Af ← SegmentRemoval(dom, C)                    ▷ procedure call
10:    end if
11:    result ← 0
12:    for all t ∈ C do
13:        result ← result + (W1 * fo(t))
14:        result ← result + (W2 * fhb(t))
15:        result ← result + (W4 * ftm(t))
16:    end for
```

(cont.)

(cont.)

```
    result ← result + (W3 * k)
    result ← result + (W5 * ft(Af)))
    result ← result * (D(C, Af))
    result ← result * (fmin(C, Af))
    OUTPUT result
end procedure
procedure SegmentRemoval(dom, C)
    Af ← dom
    NameSeg[] ← tagsSegmented()
    Af ← RemoveDisposal(Af)
    for all tagname ∈ NameSeg[] do
        for all tag ∈ getTags(Af, tagname) do
    Af ← dom
    NameSeg[] ← tagsSegmented()
    Af ← RemoveDisposal(Af)
    for all tagname ∈ NameSeg[] do
        for all tag ∈ getTags(Af, tagname) do
            if !(hasAnyKey(tag, C)) then
                Remove(Af, tag)
            end if
        end for
    end for
    OUTPUT Af
end procedure
```

Fonte:Autor

4 AVALIAÇÃO EXPERIMENTAL

Nesta seção, são descritos os experimentos realizados para demostrar a eficácia da proposta apresentada. Para análise e avaliação do processo de extração, são armazenados alguns metadados da página original e filtrada, tais como quantidade de palavras do texto, quantidade de nodos, de blocos de destaque, de blocos-segmentos, nome do arquivo, consulta realizada, código HTML além de um mapeamento de termos e nodos. Para contextualizar os resultados apresentados pelo algoritmo proposto, o modelo vetorial, implementado pelo Lucene, foi usado como baseline.

A função de ranking utiliza diferentes pesos, nestes experimentos o número de ocorrências de alguma palavra-chave tem peso $w_1 = 9{,}98$; a quantidade de palavra-chave em destaque tem peso $w_2 = 15$; a quantidade de links com todas as palavras-chaves da consulta em sua descrição, peso $w_3 = 15$; a ocorrência de cada palavra-chave no título peso $w_4 = 60$ e a quantidade de termos do documento tem peso $w_5 = 0{;}02$. Os valores utilizados de penalização das funções de correlação e *Fmin* possuem α iguais a 0.08 e 0.1 respectivamente, sendo o valor de d (definição 6) igual a 1000 caracteres e x e y (definição 7) iguais a 300 termos e 10 links respectivamente.

Os pesos foram calibrados manualmente baseado nas observações dos metadados das páginas HTML. Para calibração foram dados valores iniciais para cada critério e ajustados para mais ou para menos conforme o melhor resultado de precisão e revocação do FPR sobre as páginas HTML originais, sem a aplicação de nenhum filtro, para que só depois fosse validado se a aplicação do QBM, antes da etapa de ranking, melhoraria o ranking médio geral. Foi comum encontrar páginas HTML com mais de 10.000 ou 20.000 palavras, então um peso aparentemente insignificante para quantidade de palavras se torna significante. Depois de encontrado pesos satisfatórios (que tornam o FPR melhor que o lucene), estes foram ajustados para que o somatório dos pesos desse 1 (um).

O Lucene foi utilizado como baseline porque é amplamente utilizado em ferramentas de buscas locais com a implementação (VSM)

disponível, além de determinar a relevância do documento com relação à consulta efetuada. Lucene não tem as restrições de necessidade de reconhecimento do usuário (ClickThrough Data), do uso de Inteligência Artificial (PCR, ClickThrough Data) e a necessidade que muitas páginas dividam um mesmo template (Computing Block Importance for Searching on Web Sites). Block-Based Web Search apresenta melhores resultados se comparado ao FixedPS e usa páginas Web no método de ranking. Como trabalhos futuros é mencionado que a comparação entre FPR/QBM e Block-Based Web Search pode ser feita, com a melhoria de coletar o conteúdo textual da tag <body> ao invés da tag <title> no método Block-Based Web Search.

4.1 CONFIGURAÇÃO DAS CONSULTAS E DO CONJUNTO DE DOCUMENTOS

Páginas HTML que satisfizeram 30 consultas em 5 diferentes áreas foram coletadas do google e incluídas em uma base de páginas HTML de modo a garantir que pelo menos 10 páginas relevantes sobre cada assunto estaria contido na base de páginas HTML e a quantidade de páginas irrelevantes ficasse maior ou igual que a quantidade de páginas relevantes. Cada página recebeu um escore de 1 a 4 na seguinte escala: insignificante (1), pouco significante (2), significante (3) e muito significante (4). As páginas com escores médios de 3 ou mais foram classificadas como Relevantes e as demais como Irrelevantes.

O conjunto de documentos usado nos experimentos totaliza 1.530 páginas WEB, coletadas de diferentes sites de notícias e entretenimento ao longo do ano de 2015. As consultas realizadas para os experimentos estão associadas a cinco diferentes domínios, que servem para verificar se alguns deles se comportam muito diferentemente dos outros: história, direito, doenças, eletrônicos e personalidades. A tabela 4 apresenta a configuração, em números, para cada um dos domínios. A coluna PR-All indica a quantidade de páginas relevantes nas quais constam todos os termos da busca usados na consulta.

Tabela 4 - Consultas realizadas para os experimentos

Domínio	Termos de Busca	PR-All
História	Segunda Guerra Mundial	11
	História Ceará	11
	Descobrimento Brasil	14
	Primeira Guerra Mundial	13
	História Estado Santa Catarina	11
	Independência do Brasil	11
Direito	Serviço Público Conceito	13
	Direitos Fundamentais	10
	Ação Direta de Inconstitucionalidade	11
	Democracia	10
Doenças	Febre Aftosa	13
	Ebola	14
	Autismo	14
	Tuberculose	12
	Dengue	14
Eletrônicos	TV LED	11
	Tablet	22
	Impressora 3D	10
	IPhone	23
Personalidades	Dilma Rousseff	15
	Marco Feliciano	10
	Guarrincha	10
	Ronaldinho Gaúcho	11
	Dom Pedro I	10
	Pelé	10

Fonte:Autor.

4.2 METODOLOGIA E MÉTRICAS DE AVALIAÇÃO

Para fins de comparação com o algoritmo proposto FPR, os experimentos foram configurados utilizando combinações de algoritmos de *ranking* e de *extração de ruídos*. Como baseline de algoritmo de ranking, foi escolhido o modelo vetorial (através da implementação

disponibilizada pelo Lucene) e, para fins de extração de conteúdo irrelevante, o algoritmo CETR. Os testes foram conduzidos levando-se em consideração as seguintes configurações:

a. **Lucene**: algoritmo de ranking do modelo vetorial clássico, usando a implementação do Lucene sobre a base de documentos sem aplicação de filtro sobre eles;
b. **FPR(-)**: o algoritmo de ranking proposto neste trabalho sobre a base de documentos sem aplicação de filtro sobre eles;
c. **FPR + CETR**: algoritmo de ranking proposto neste trabalho, sobre a base de documentos filtrados através do algoritmo CETR;
d. **FPR + QBM**: algoritmo de ranking proposto neste trabalho, sobre a base de documentos filtrados através do algoritmo QBM.

O foco dos algoritmos propostos está na qualidade dos resultados retornados. Portanto, as métricas usadas para avaliação foram as clássicas da comunidade de recuperação de informação (BAEZA-YATES et al. 1999): revocação, precisão e medida F. O valor de revocação foi obtido através da proporção de documentos relevantes, que, em cada consulta, de fato foram recuperados. A precisão foi calculada pela proporção dos recuperados considerados relevantes. A medida-F é a média harmônica entre revocação e precisão.

Os experimentos descritos nesta seção avaliam o algoritmo FPR proposto. Primeiro, é analisada a efetividade em eliminar ruídos do método QBM e, então, é feita uma avaliação do FPR com ou sem o uso de filtros.

4.3 ANÁLISE DO QBM

O intuito do QBM é preservar áreas contínuas de texto próximas das palavras-chave da consulta, delimitadas por nodos específicos no qual se considera que não haverá quebra de ligação entre as palavras, eliminando os nodos considerados blocos-descarte e blocos-segmentos que não possuem palavras-chave da consulta nele e em nenhum de seus nós descendentes.

O algoritmo QBM foi analisado em sua efetividade para

remover ruídos de páginas HTML em comparação com o baseline CETR (WENINGER et al., 2010). Para avaliar os resultados, os dois algoritmos foram testados num mesmo conjunto de páginas. Os parâmetros considerados foram:

(i) *qtdPUDF*: total de palavras úteis da DOM filtrada;
(ii) *qtdPDF*: total de palavras da DOM filtrada;
(iii) *qtdPUDO*: total de palavras úteis da DOM original.

De posse destes parâmetros, foi possível avaliar a precisão e revocação da seguinte forma: *revocação = qtdPUDF/qtdPUDO*; *precisão = qtdPUDF/qtdPDF*.

Os testes foram realizados em páginas aleatórias e calculados em termos de *precisão*, *revocação* e *f-value*. O QBM alcançou mais de 80% de precisão, sendo melhor que o CETR na maioria dos casos, conforme podemos verificar na tabela 5, sendo, portanto, mais indicado para uso com o FPR, conforme seu propósito de ser um método de ranking de páginas HTML que tiveram os conteúdos irrelevantes excluídos.

Tabela 5 - Revocação e Precisão - QBM x CETR

Página	**Total de Palavras**	***qtdPUDO***	***qtdPDF***		***qtdPUDF***		**Revocação**		**Precisão**		**F-Measure**	
			QBM	**CETR**	**QBM**	**CETR**	**QBM**	**CETR**	**QBM**	**CETR**	**QBM**	**CETR**
1	2223	1964	1759	1631	1759	1538	0.896	0.783	1.000	0.943	0.948	0.856
2	618	163	464	442	160	145	0.982	0.890	0.345	0.328	0.510	0.479
3	1078	738	811	28	733	0	0.993	0	0.904	0	0.946	0
4	5879	5108	3339	1912	3291	1841	0.644	0.360	0.986	0.963	0.815	0.525

(cont.)

(continuação)

5	2816	1855	1826	1836	1793	1821	0.967	0.982	0.982	0.992	0.975	0.987
6	623	328	328	322	328	322	1.000	0.982	1.000	1.000	1.000	0.991
7	1207	389	288	703	288	348	0.740	0.895	1.000	0.495	0.87	0.637
8	1311	0	868	1023	0	0	0	0	0	0	0	0
9	703	348	314	374	293	328	0.842	0.943	0.933	0.877	0.885	0.909
10	1722	1308	1271	1229	1270	1189	0.971	0.909	0.999	0.967	0.985	0.937
-	-	-	-	-		Média	0.803	0.674	0.815	0.657	0.787	0.632

Fonte:Autor.

A tabela 5 apresenta o resultado de 10 páginas analisadas (uma lista de um conjunto analisado para que possamos abordar alguns pontos). A página listada com número 8 tem 0% de precisão e revocação devido ao fato de tanto o QBM quanto o CETR não avaliarem o objeto relacionado aos termos da busca, o que demandaria um trabalho muito árduo de inteligência artificial e, consequentemente, maior complexidade do algoritmo QBM, que é da ordem O(n). Por exemplo, numa situação em que os termos da consulta são "*história*" e "*ceará*" e o objeto do assunto pesquisado seja "história do Estado do Ceará", o filtro pode extrair a região considerada útil de uma página com conteúdo relacionado ao objeto "história do time Ceará Esporte Clube", que é uma página inútil para o objeto de "história do Estado do Ceará", resultando em revocação e precisão 0% nos casos específicos em que as mesmas palavras são utilizadas para objetos de pesquisas distintos.

O mesmo acontece com a página 2, no qual uma página pouco relevante combina com a consulta "serviço público conceito" e no mesmo bloco-segmento traz um texto sobre "servidor público" e

somente uma pequena parte aborda o conceito de serviço público. Na página 3, o CETR extrai os nós da árvore DOM, formando uma árvore filtrada, mas somente com conteúdo irrelevante para a consulta. A medida f-measure é a mais indicada para avaliar a efetividade do filtro, pois não adianta um filtro que tenha alta precisão e baixa revocação, nem o inverso.

Para avaliar os filtros, extraímos o código HTML das páginas filtradas e exibimos em um navegador o resultado. Como as páginas Web são salvas somente pelo conteúdo HTML, sem incluir arquivos anexos, elas podem aparecer com layout desformatado no navegador, mas isto não afeta o resultado final, pois, para o FPR, só importa o texto extraído dos nós HTML e, quando apresentadas em ordem decrescente de ranking, somente as páginas Web originais são exibidas aos usuários.

Para uma pequena demonstração e melhor entendimento do funcionamento dos filtros, anexamos 3 páginas Web originais (Anexos 1, 4 e 7) e o resultado da aplicação dos filtros QBM e CETR sobre elas.

O anexo 1 trata de uma página Web que apresenta informações variadas sobre tablets e realiza o comparativo em tablets de 4 marcas (Sony, Samsumg, Apple e Motorola); nesta página, há três grandes regiões localizadas no cabeçalho, no centro e no rodapé e somente a parte central apresenta conteúdo relevante para a consulta efetuada.

Podemos perceber, pelo anexo 2, que o filtro QBM eliminou praticamente todo conteúdo irrelevante do cabeçalho e do rodapé, sendo, neste caso, uma parte maior de conteúdo irrelevante preservada pelo CETR, conforme podemos perceber no Anexo 3. Além disto, nem todo conteúdo útil foi preservado, tanto pelo QBM quanto pelo CETR, sendo mais conteúdo útil aproveitado pelo QBM.

No anexo 4, temos uma página Web que trata da Segunda Guerra Mundial. Pelos anexos 5 e 6, percebemos que tanto o QBM quanto o CETR conseguiram aproveitar todo conteúdo relevante, sendo praticamente todo conteúdo irrelevante extraído pelo QBM (exceção da sessão “comente”, na qual restou uma palavra) e também extraído pelo CETR, que só preservou uma pequena parte do conteúdo irrelevante mostrado pelo anexo 6 (em vermelho).

Os anexos 7, 8 e 9 demonstram que uma página original sobre

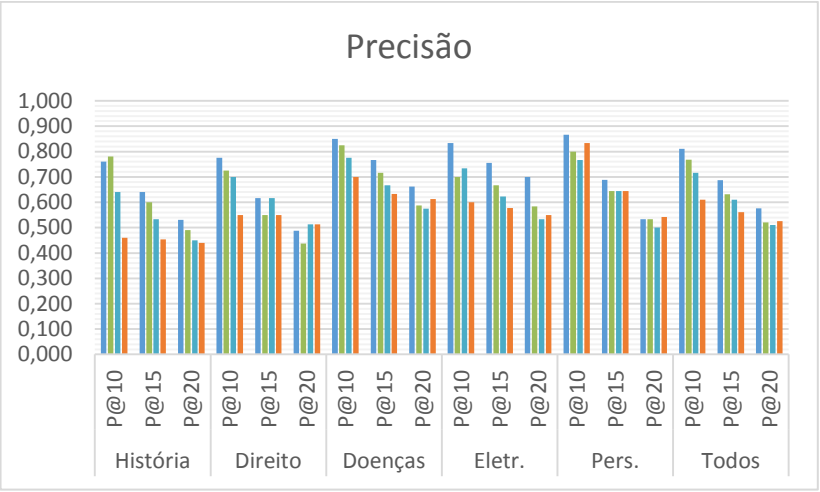
tv led teve quase todo conteúdo irrelevante (representado pela área delimitada em vermelho no Anexo 7) extraído pelo QBM, restando somente dois componentes de pesquisa interna como conteúdo irrelevante preservados e que o CETR apesar de aproveitar todo conteúdo relevante, também trouxe muito conteúdo irrelevante (ocultos ou não) presentes no cabeçalho e rodapé.

4.4 ANÁLISE DO FPR

Os resultados da aplicação do algoritmo FPR são descritos em dois grandes grupos, considerando as análises realizadas.

No primeiro grupo, cada domínio foi analisado individualmente, de forma a identificar se algum deles se comportaria de forma muito diferente do geral. No segundo grupo, são apresentados os resultados gerais, considerando valores de média sobre todo o conjunto de documentos (os 1.530 documentos), de forma a ter uma ideia geral do comportamento, independentemente da distribuição dos dados nos domínios.

Figura 5 - Precisão P@10, P@15, P@20

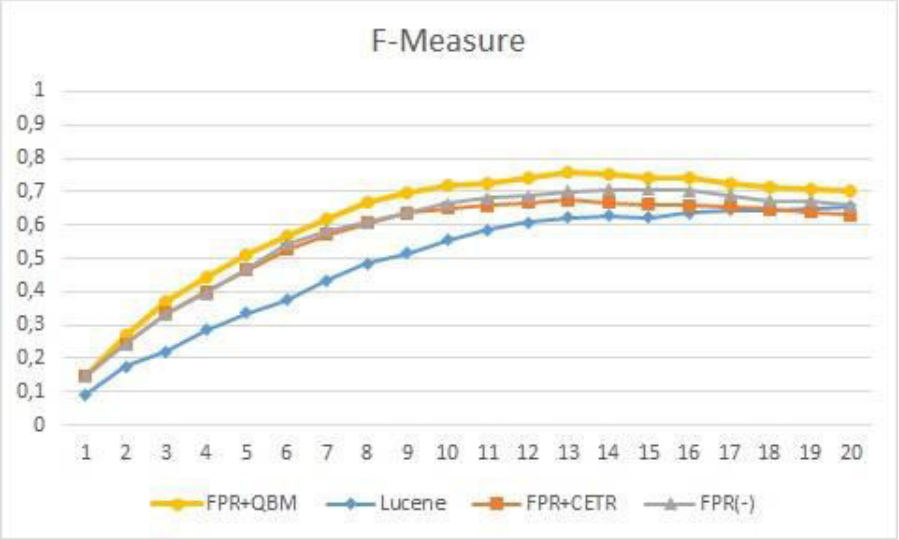


Fonte: Autor.

A figura 5 apresenta os resultados de precisão com experimentos realizados em três diferentes rankings, conforme medida clássica da área de recuperação de informação.

O primeiro ranking, até a posição 10; o segundo, até a posição 15 e o terceiro, até a posição 20. Da esquerda para direita, cada grupo de 4 barras verticais representa **FPR + QBM**, **FPR(-)**, **FPR+CETR e Lucene,** respectivamente. Pela análise da figura 5, pode-se perceber que FPR+QBM possui precisão média melhor que o baseline (Lucene). O filtro QBM fez com que a precisão média do FPR melhorasse nos ranking P@10, P@15 e P@20 e apresentou melhores resultados que o baseline CETR (método automático de extração da região principal de uma página).

Figura 6 - F-Measure 20 primeiras posições



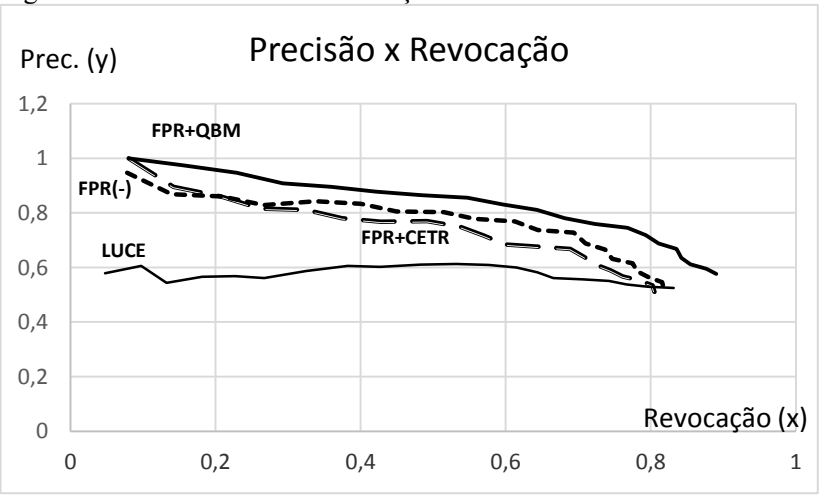
Fonte: Autor

Os resultados da avaliação de valores de F-Measure (eixo y) nas 20 primeiras posições (eixo x) são apresentados na figura 6, levando em consideração a precisão e a revocação média das consultas realizadas. Como constatado, o FPR funciona melhor com o QBM se comparado ao

baseline CETR. Quando comparamos a medida clássica de *Revocação x Precisão,* o FPR+QBM oferece o melhor resultado nas pesquisas, mostrando que o método criado para ranking de páginas filtradas e o filtro QBM atingem bons resultados, se comparados ao baseline.

Pela análise dos dados, nota-se que a média da revocação, da precisão e da medida f-measure nas primeiras 10 posições se dão melhor com a aplicação do método proposto do que com o uso do Lucene, que utiliza a técnica VSM para definir quão próximo um texto é de uma pesquisa efetuada. Além disto, o FPR apresenta melhores resultados quando utilizado em conjunto com o QBM. Apesar de o filtro ser específico para determinadas tags HTML, ele funciona com boa precisão (acima de 85%) para a maioria das páginas testadas, que utilizam de tags *div* e *span* para posicionar e agrupar conteúdo e não apenas tabelas, mantendo a precisão do filtro em páginas de e-commerce que usam amplamente estas tags para apresentação de conteúdos.

Figura 7 - Curva Precisão x Revocação



Fonte: Autor

CONCLUSÕES E TRABALHOS FUTUROS

Este trabalho apresentou um processo de ranking de páginas HTML que avalia quão próximo um documento é de uma consulta com base em critérios (termos em destaque, frequência dos termos, correlação dos termos...) considerados em estudo de recuperação da informação aplicados ao contexto de páginas Web. Também foi verificado que o FPR, aplicado em páginas que tiveram seu conteúdo irrelevante para consulta eliminado, melhora a qualidade do ranking do FPR. O processo de filtro envolve segmentação de páginas HTML e definição do conteúdo relevante para consulta efetuada através da identificação de blocos-segmentos que estão ligados à consulta e com isto extração de conteúdo irrelevante da árvore DOM filtrada que será avaliada na etapa de definição da relevância, contendo somente blocos-segmentos ligados à consulta; assim, os documentos são segmentados em blocos. São excluídos os blocos considerados irrelevantes para a consulta efetuada, calculando-se em seguida quão relevante um documento é para a consulta do usuário. O método de ranking proposto, chamado Filtered-Page Ranking, possui duas etapas principais:

(i) segmentação da página Web e eliminação do conteúdo irrelevante para consulta;
(ii) ranking dos documentos.

Pode-se concluir que o FPR(com aplicação do QBM) é um método de ranking, com prévia eliminação de conteúdo irrelevante, satisfatório se comparado com métodos da literatura que podem ser usados para definir a relevância de páginas HTML em relação à busca efetuada, além de servir como um método de ranking que possa ser utilizado e melhorado.

Não é o foco do trabalho proposto melhorar o ranking do Lucene utilizando o modelo vetorial, os experimentos iniciais demonstraram que tanto o QBM quanto o CETR pioram o ranking do Lucene, devido a própria natureza do modelo vetorial. Quanto mais irrelevante for a página Web para determinada consulta, menos conteúdo tende a ser preservado após esta passar pelo filtro QBM. Quanto menos conteúdo texto em documentos que possuam as palavras-chaves da

consulta, melhor este é avaliado pelo método VSM, recebendo maior escore, por ser considerado um documento com conteúdo mais próximo da consulta efetuada.

Propomos como trabalhos futuros, avaliar se o QBM melhora a precisão de outros métodos de ranking específicos de páginas Web que avaliem quão próximo uma página Web é de uma consulta efetuada, como o Block-Based Web Search, com a melhoria de coletar o conteúdo textual da tag <body> ao invés da tag <title> no método Block-Based Web Search.

Com relação ao método FPR, acreditamos que futuramente poderão ser feitos trabalhos como encontrar um nível ótimo de pesos dos critérios relevantes para a definição do ranking, desenvolver um processo de indexação que dê suporte aos dados e metadados interligados com o método de ranking FPR, de tal modo que permita a criação de uma máquina de busca completa (análise, indexação, pesquisa e ranking) em uma base local de documentos HTML, além de fornecer novos critérios relevantes para definição de ranking como *inverse document frequence* (IDF) e explorar novos critérios relevantes como a presença de vídeos nas páginas HTML.

REFERÊNCIAS

ADALI, S.; HILL, B. and MAGDON-ISMAIL, M. The impact of ranker quality on rank aggregation algorithms: Information vs. robustness. In *Data Engineering Workshops,* 2006. Proceedings. 22nd International Conference on . IEEE, pp. 37–37, 2006.

BAEZA-YATES, R., RIBEIRO-NETO, B. et al. *Modern information retrieval* . v. 463. ACM press New York, 1999.

BALOG, K. and RAMAMPIARO, H. Cumulative citation recommendation: Classification vs. ranking. In Proceedings of the 36th International ACM SIGIR Conference on Research and Development. In *Information Retrieval*. SIGIR'13.

BORGES, G. S. B. *Indexação Automática de Documentos Textuais*: Proposta de Critérios Essenciais. Universidade Federal de Minas Gerais, Brazil (UFMG – ECI), Brazil, 2009.

BURGET, R. and RUDOLFOVA, I. WEB page element classification based on visual features. In Intelligent Information and Database Systems, 2009. ACIIDS 2009. *First Asian Conference on IEEE*, 2009, p. 67–72.

CAI, D. , YU, S. , WEN, J.-R. , and MA, W.-Y. Extracting content structure for WEB pages based on visual representation. In WEB Technologies and Applications. Springer, pp. 406–417, 2003.

CAI, Deng et al. Block-based web search. In: Proceedings of the 27th annual international ACM SIGIR conference on Research and development in information retrieval. ACM, 2004. p. 456-463.

CALLAN, J. P. Passage-level evidence in document retrieval. In *Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval.* Springer-Verlag New York, Inc., 1994, p. 302–310.

CRESCENZI, V. et al. Roadrunner: Towards automatic data extraction from large WEB sites. In *VLDB*. v. 1. 2001. p. 109–118.

DEVI, P. , GUPTA, A. , and DIXIT, A. Comparative study of hits and pagerank link based ranking algorithms. *International Journal of Advanced Research in Computer and Communication Engineering* 3 (2), 2014.

DUHAN, N. , SHARMA, A. , and BHATIA, K. K. Page ranking algorithms: a survey. In *Advance Computing Conference*, 2009. IACC 2009. IEEE International. IEEE, pp. 1530–1537, 2009.

Dwork, C. et al. *Rank aggregation revisited,* 2001.

Fang, J. et al. Ontology-based automatic classification and ranking for WEB documents. In *Proceedings of the Fourth International Conference on Fuzzy Systems and Knowledge Discovery* – v. 03

FSKD ’07. *IEEE Computer Society*, Washington, DC, USA, 2007. p. 627–631.

Fernandes, D et al. Computing block importance for searching on WEB sites. In *Proceedings of the sixteenth ACM conference on Conference on information and knowledge management.* ACM2007, p. 165–174.

Hatcher, E. , Gospodnetic, O. , and McCandless, M. *Lucene in action,* 2004.

INSA, D.; SILVA, J.and TAMARIT, S. *Using the words/leafs ratio in the dom tree for content extraction. J. Log. Algebr. Program*. 2013.

JOACHIMS, T. Optimizing search engines using clickthrough data. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining* . ACM, 2002. p. 133–142.

KARATZOGLOU, A.; BALTRUNAS, L. and SHI, Y. Learning to rank for recommender systems. In *Proceedings of the 7th ACM Conference on Recommender Systems.* RecSys ’13. ACM, New York, NY, USA, 2013.p. 493–494.

KISE, K.; SATO, A. and IWATA, M. *Segmentation of page images using the area voronoi diagram. Computer Vision and Image Understanding.* 1998.

KLEINBERG, J. M. Authoritative sources in a hyperlinked environment. *Journal of the ACM* (JACM).1999.

LANGVILLE, A. N. and MEYE, C. D. *Google’s PageRank and Beyond*: The Science of Search Engine Rankings. Princeton Ed., 2011.

LERCHE, L. and JANNACH, D. Using graded implicit feedback for bayesian personalized ranking. In *Proceedings of the 8th ACM Conference on Recommender Systems* . RecSys '14. ACM, New York, NY, USA, pp. 353–356, 2014.

LI, J.; SAHA, B. and DESHPANDE, A. A unified approach to ranking in probabilistic databases. *The VLDB Journal* . Apr., 2011.

LIU, B. In *Structured Data Extraction: Wrapper Generation* . Springer, 2011. p. 396–406.

MOLKOVÁ, L. *Indexing very large text data*. Brno, spring, 2011.

Page, L. et al. *The pagerank citation ranking:* bringing order to the WEB. 1999.

POKORNY, J. and SMIZANSKY, J. Page content rank: an approach to the WEB content mining. In *Proceedings of IADIS International Conf. On Applied Computing* . v. 1. 2005. p. 289–296,

SALTON, G. and BUCKLEY, C. Term-weighting approaches in automatic text retrieval. *Inf. Process. Manage.* Aug., 1988.

SANCHES, P. A. G. *Aprendendo funções de ranking baseadas em blocos usando programação genética*. Universidade Federal do Amazonas, Brazil (UFAM), Brazil, 2013.

SELVAN, M. P. , SEKAR, A. C. , and DHARSHINI, A. P. Survey on WEB page ranking algorithms. *International Journal of Computer Applications*. 2012.

SONG, Y.; ZHANG, L. and GILES, C. L. Automatic tag recommendation algorithms for social recommender systems. *ACM Trans.* WEB. Feb., 2011.

SU, A.J. et al. How to improve your search engine ranking: Myths and reality. *ACM Trans. WEB*. Mar., 2014.

VELLOSO, R. P. and DORNELES, C. F. Automatic WEB page segmentation and noise removal for structured extraction using tag path sequences. *Journal of Information and Data Management.* 2013.

WANG, J. and LOCHOVSKY, F. H. Data-rich section extraction from html pages. In WEB Information Systems Engineering, 2002. WISE 2002. *Proceedings of the Third International Conference on . IEEE,* 2002. p. 313–322.

WENINGER, T.; HSU, W. H. and HAN, J. Cetr: content extraction via tag ratios. In *Proceedings of the 19th international conference on World wide WEB* . ACM, pp. 2010. p. 971–980.

XI, Wensi et al. "Incorporating window-based passage-level evidence in document retrieval." *Journal of information science* 27.2. 2001. p. 73-80.

ZHU, H. et al. Ranking fraud detection for mobile apps: A holistic view. In Proceedings of the 22Nd ACM International Conference on Conference on *Information & Knowledge Management*. CIKM'13. ACM, New York, NY, USA, pp. 619–628, 2013.

ANEXOS

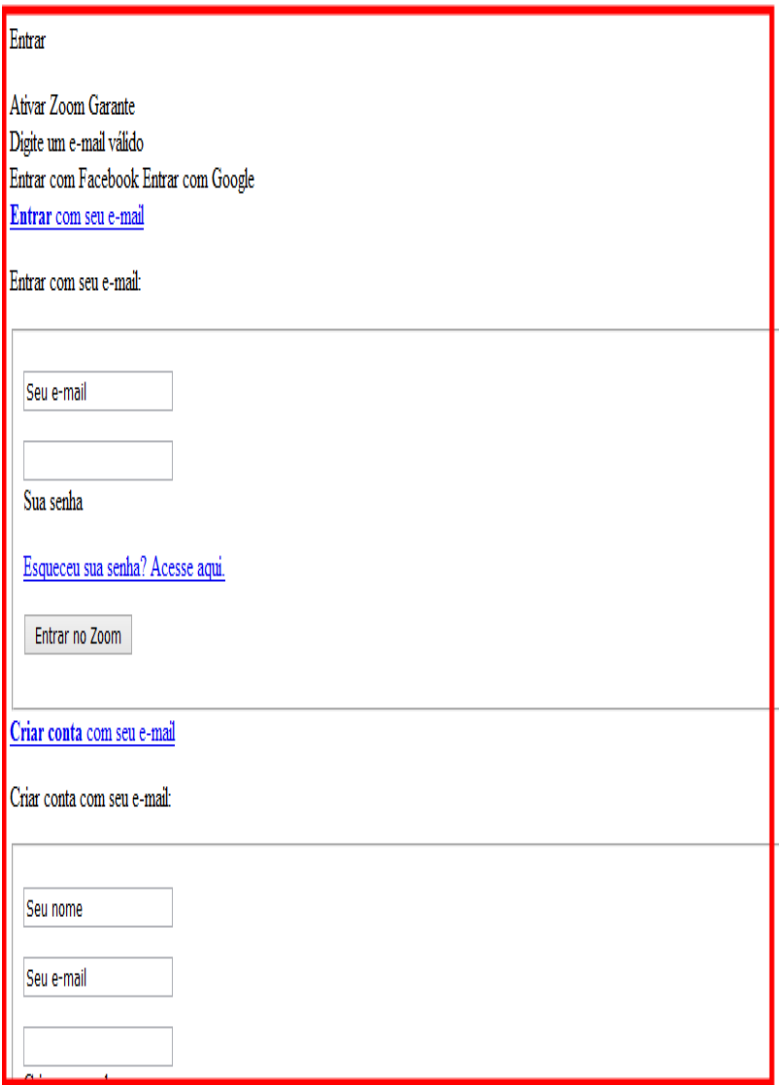
Anexo 1 – Página Original (Consulta: Tablet)

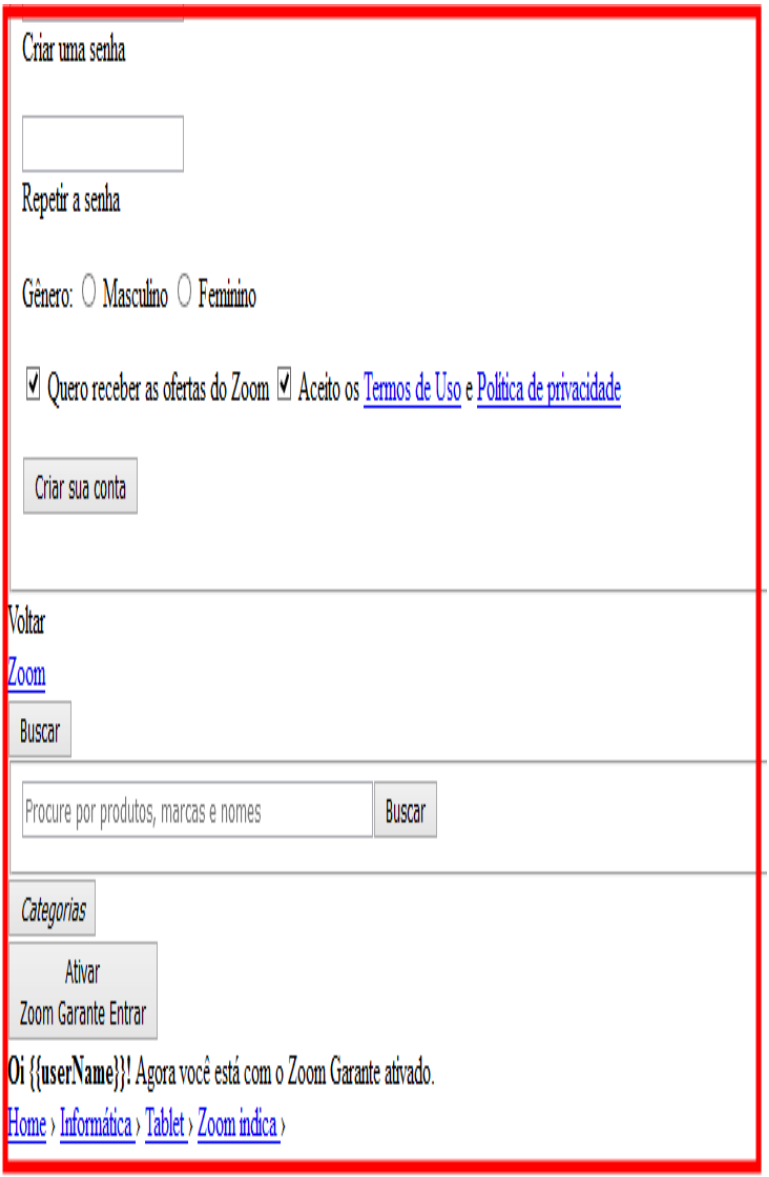
- publicidade
- Eletrodomésticos
  - Ar Condicionado
  - Lavadora de Roupas
  - Fogão
  - Geladeira
  - Microondas
  - Forno
  - Mais Eletrodomésticos
  - publicidade
- Câmeras e Filmadoras
  - Câmera Digital
  - Filmadora
  - Lente para Câmera
  - Bolsa e Capa para Filmadora
  - Tripé
  - Outros Acessórios para Câmera
  - Mochila para Câmera
  - Mais Câmeras e Filmadoras
  - publicidade
- Móveis
  - Sofá
  - Guarda-Roupas
  - Cama
  - Rack, Painel e Estante para TV
  - Armário e Balcão de Cozinha
  - Mesas e Cadeiras para Sala de Jantar
  - Mais Móveis
  - publicidade
- Esporte e Lazer

- Tênis
- Bicicleta
- Chuteiras
- Camisas de Times de Futebol
- Esteira Ergométrica
- Monitor Cardíaco
- Mais Esporte e Lazer
- publicidade

- Moda e Acessórios
  - Tênis
  - Sapato
  - Camisa, Camiseta e Blusa
  - Sandália
  - Bota
  - Boné
  - Mais Moda e Acessórios
  - publicidade
- Todas as Categorias
  1. Automóveis e Veículos
  2. Bebês
  3. Beleza e Saúde
  4. Brinquedos
  5. Cama, Mesa e Banho
  6. Casa e Jardim
  7. CD, DVD e Blu-Ray
  8. Celulares e Telefones
  9. Câmeras e Filmadoras
  10. Decoração de Natal
  11. Eletrodomésticos


1. Eletroportáteis
2. Esporte e Lazer
3. Flores e Cestas
4. Games
5. Indústria e Comércio
6. Informática
7. Instrumentos Musicais
8. Joalheria
9. Livros e Revistas
10. Moda e Acessórios
11. Móveis e Decoração
1. Papelaria e Armarinho
2. Perfumaria
3. Pet Shop
4. Sex Shop
5. Supermercado Online
6. Tabacaria
7. TV e Eletrônicos
8. Utilidades Domésticas
9. Todas as Categorias
Zoom Garante
Se a loja não entregar, a gente resolve.
Só quem tem as lojas mais confiáveis pode garantir: se o produto não chegar, resolvemos direto com a loja ou devolvemos seu dinheiro em até R$3.000. Conheça o serviço. É grátis
Importante: Você precisa estar identificado no Zoom antes de acessar a loja no dia da compra.
Fechar


Entrar
Ativar Zoom Garante
Digite um e-mail válido
Entrar com Facebook Entrar com Google
Entrar com seu e-mail
Entrar com seu e-mail:
Seu e-mail
Sua senha
Esqueceu sua senha? Acesse aqui.
Entrar no Zoom
Criar conta com seu e-mail
Criar conta com seu e-mail:
Seu nome
Seu e-mail


Criar uma senha
Repetir a senha
Gênero: Masculino Feminino
Quero receber as ofertas do Zoom Aceito os Termos de Uso e Política de privacidade
Criar sua conta
Voltar
Zoom
Buscar
Procure por produtos, marcas e nomes
Buscar
Categorias
Ativar
Zoom Garante Entrar
Oi {{userName}}! Agora você está com o Zoom Garante ativado.
Home › Informática › Tablet › Zoom indica ›

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Zoom indica em Tablet

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Sony S, Samsung Galaxy Tab 10.1, iPad 3 ou Motorola Xoom 2? Saiba qual tablet sai na frente.

Artigo atualizado em 16/01/2015

Gostou? Compartilhe!

- Facebook
- twitter
- Google

por **Ricardo Bergher** - Especialista em Tecnologia.

Na hora de comprar um tablet, com tantas boas opções disponíveis no mercado, você fica até em dúvida na hora de escolher o modelo ideal para você? Sim, pois modelos não faltam, e muitas vezes ficamos até perdidos na hora de decidir o melhor tablet para nossas tarefas do dia a dia.

Contudo, é sempre bom lembrar que a primeira coisa que você precisa fazer é perceber qual uso você vai dar para o tablet. Ele será só para entrar nas redes sociais? Para ver filmes? Ou mais para jogar? Com essas questões em mente fica mais fácil saber qual será o tablet com a configuração certa para suas atividades.

Para ajudar você, selecionamos alguns tablets de marcas bem conhecidas para uma disputa: o tablet Sony S, o Samsung Galaxy Tab 10.1, o Novo iPad (ou iPad 3) e o Motorola Xoom 2. Confira o que cada modelo oferece de mais interessante e faça a escolha certa.

Sistema Operacional
Tamanho da Tela
Peso
Conectividade
GPS
Bluetooth
Memória Interna
Câmera Frontal
Câmera Principal
Filma em HD

	Sony S	Galaxy Tab 10.1	iPad 3	Motorola Xoom 2
				
Sistema Operacional	Android 4.0	Android 3.1	iOS 5	Android 3.2
Tamanho da Tela	9.4"	10.1"	9.7"	10.1"
Peso	598g	565g	662g	603g
Conectividade	Wi-Fi	Wi-Fi, 3G e 4G	Wi-Fi, 3G e 4G	Wi-Fi e 3G
GPS	Sim	Sim	Sim	Sim
Bluetooth	Sim	Sim	Sim	Sim
Memória	32 GB	16 GB	16 / 32 / 64 GB	32 GB

Câmera Frontal	0.3 MP	2 MP	0.3 MP	1.3 MP
Câmera Principal	5 MP	3 MP	5 MP	5 MP
Filma em HD	Sim	Sim	Sim	Sim
Adobe Flash	Sim	Sim	Não	Sim

Depois de conhecer as principais configurações desses tablets, como conexões de internet, câmeras e sistemas operacionais, fica mais prático saber o modelo que entrega os recursos que você tanto precisa. E o melhor de tudo é saber que aqui no Zoom você encontra muitos outros tablets, sendo fácil descobrir o melhor modelo para suas tarefas.

Outros modelos de tablets para você

- Foto Tablet Samsung Galaxy Tab 3 TV 3G 8 GB TFT 7" Android 4.1 (Jelly Bean) 3 MP SM-T211M

  Tablet Samsung Galaxy Tab 3 TV 3G 8 GB TFT 7" Android 4.1 (Jelly Bean) 3 MP SM-T211M

  a partir de: R$ 1.406,14 em 2 lojas

  Veja mais
- Foto Tablet Apple iPad Mini 16 GB 7.9" Wi-Fi 5 MP

  Tablet Apple iPad Mini 16 GB 7.9" Wi-Fi 5 MP

  a partir de: R$ 891,03 em 4 lojas

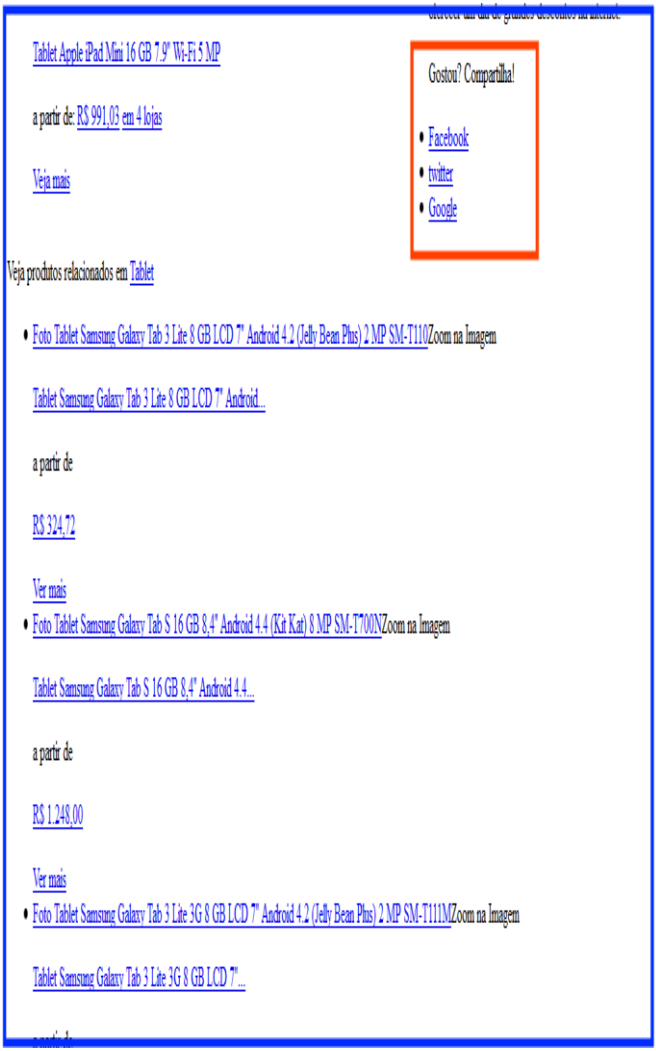
Apple e Samsung, por exemplo, que já são marcas conhecidas por entregarem smartphones de sucesso, mantêm o mesmo caminho no mundo dos tablets, oferecendo, geração após geração, modelos cada vez mais modernos.

Então, se você quiser conhecer um tablet Samsung ou um iPad, no Zoom tem vários modelos para você escolher, e até mesmo de outras marcas. Aproveite!

Ah, e se você é daqueles que adora fazer pesquisas antes de comprar qualquer produto, aproveite para comparar preços. Assim você conhece as melhores ofertas e faz a escolha certa! Aliás, não custa lembrar: aqui você encontra durante todo o ano muitas ofertas, e não apenas no dia da Black Friday, data já conhecida por oferecer um dia de grandes descontos na internet.

Gostou? Compartilhe!

- Facebook



Tablet Apple iPad Mini 16 GB 7.9" Wi-Fi 5 MP

a partir de: R$ 991,03 em 4 lojas

Veja mais

Gostou? Compartilha!

- Facebook
- twitter
- Google

Veja produtos relacionados em Tablet

- Foto Tablet Samsung Galaxy Tab 3 Lite 8 GB LCD 7" Android 4.2 (Jelly Bean Plus) 2 MP SM-T110Zoom na Imagem

  Tablet Samsung Galaxy Tab 3 Lite 8 GB LCD 7" Android...

  a partir de

  R$ 324,72

  Ver mais
- Foto Tablet Samsung Galaxy Tab S 16 GB 8,4" Android 4.4 (Kit Kat) 8 MP SM-T700NZoom na Imagem

  Tablet Samsung Galaxy Tab S 16 GB 8,4" Android 4.4...

  a partir de

  R$ 1.248,00

  Ver mais
- Foto Tablet Samsung Galaxy Tab 3 Lite 3G 8 GB LCD 7" Android 4.2 (Jelly Bean Plus) 2 MP SM-T111MZoom na Imagem

  Tablet Samsung Galaxy Tab 3 Lite 3G 8 GB LCD 7"...

a partir de

R$ 424,15

Ver mais

Top 5 em tablet

+ em Tablet

Guia de compras Tá na dúvida? Dê um Zoom em Tablet

Entenda antes de comprar

Não sabe por onde começar? Aqui você encontra todas as informações necessárias para entender tudo sobre recursos e tecnologia.

- O que posso fazer com um tablet?

  Descubra o que faz um tablet e aproveite todos os seus recursos!

- Por que ter um Tablet?

  Saiba os diferenciais que você vai ter ao comprar um tablet.

- Tablet Samsung Galaxy Tab 3 Lite SM-T110 8 GB Android 7

  Tablet Samsung Galaxy Tab 3 Lite SM-T110 8 GB Android 7"

  19 Avaliações

  a partir de

  R$ 324,72 em 11 lojas

  Kalunga R$ 449,00 Ir à loja

  Magazine Luiza R$ 349,00 Ir à loja

  Ricardo Eletro R$ 449,10 Ir à loja

- Tablet LG G Pad V700 16 GB Android 10.1

  Tablet LG G Pad V700 16 GB Android 10,1" 5 MP

  11 Avaliações

- 

  Tablet ou notebook: qual escolher?

  Descubra o ideal para suas tarefas: notebook ou tablet?

- 

  Qual a memória interna ideal para o seu tablet?

  Descubra qual memória interna do tablet não vai deixar você na mão.

Ver todas as matérias

Zoom indica

Pensando naquilo que você precisa, selecionamos várias dicas de marcas e produtos para ajudá-lo a fazer a melhor compra.

- 

  Tudo sobre os tablets Samsung Galaxy

  Com tantas opções, veja qual tablet Samsung é ideal para o seu dia a dia.

- 

  O que fazer para a bateria do tablet Android durar mais?

  Veja dicas para melhorar a duração da bateria do tablet Android.

- 

a partir de

R$ 719,10 em 8 lojas

Fnac R$ 899,10 Ir à loja

Magazine Luiza R$ 782,13 Ir à loja

Americanas R$ 719,10 Ir à loja

- Tablet HP 1201 8 GB Android 7.1

  Tablet HP 1201 8 GB Android 7,1" 2 MP 4.2 (Jelly Bean Phis)

  4 Avaliações

a partir de

R$ 269,10 em 5 lojas

Kalunga R$ 349,00 Ir à loja

Americanas R$ 269,10 Ir à loja

Submarino R$ 269,10 Ir à loja

- Tablet Asus Fonepad 7 ME372CG 8 GB 3G Android

  Tablet Asus Fonepad 7 ME372CG 8 GB 3G Android

  24 Avaliações


Conheça os tablets com celular
Um tablet com celular ajuda na comunicação. Confira alguns modelos.
iPad Air 2 ou iPad Mini 3: qual comprar?
Descubra as diferenças entre o iPad Air 2 e o iPad Mini 3 e escolha o seu!
Ver todas as matérias
a partir de
R$ 521,13 em 2 lojas
Kalunga R$ 699,00 Ir à loja
Magazine Luiza R$ 521,13 Ir à loja
Tablet LG G Pad V400 8 GB Android 7
Tablet LG G Pad V400 8 GB Android 7" 3 MP
12 Avaliações
Guia de compras de outros produtos
Anterior
Próxima
Imagina só receber as melhores ofertas de bandeja por e-mail. Bom, né? Política de privacidade
É só colocar seu e-mail ac
Gênero: Masculino Feminino
Manda pra mim
a partir de
R$ 485,19 em 6 lojas
Fnac R$ 539,10 Ir à loja
Americanas R$ 539,10 Ir à loja
TV
Notebook
Celular
Câmera Digital
Tablet
Geladeira
Lavadora
PC / Computador
Console de Videogame
Fogão
Ar Condicionado
GPS
Zoom Somos um comparador de preços e produtos feito por pessoas que só pensam numa coisa: ajudar você a fazer a melhor compra, sempre. Por isso, dê um Zoom e conte com a

Zoom Somos um comparador de preços e produtos feito por pessoas que só pensam numa coisa: ajudar você a fazer a melhor compra, sempre. Por isso, dê um Zoom e conte com a gente. :)

O Zoom

- Sobre o Zoom
- Fale com a gente
- Trabalhe com a gente
- FAQs
- Guia do Consumidor

Zoom Garante

- Sobre o Zoom Garante
- A loja não entregou o produto? Abrir reclamação

Marcas e Lojas

- Anuncie com a gente
- Área do anunciante

Hotsites

- Especiais
- Black Friday

eu Zoom:

- facebook
- twitter
- youtube
- google plus
- Aplicativo do Zoom na Play Store
- Aplicativo do Zoom na Apple Store

- GPS
- Cafeteira
- MP3 e MP4 Player
- DVD Player Automotivo
- Blu-Ray

Aplicativo do Zoom na Apple Store

O uso deste site está sujeito aos termos e condições do Termo de Uso e Política de privacidade. © Zoom. Todos os direitos reservados.

Zoom garanteX

Quer garantir a sua entrega?

É grátis! Basta fazer seu login.

Não garantirOK, fazer login

ok

X

Rapidinho... antes de ir à loja, ative o Zoom Garante! É de graça.O que é ?

Ir à loja com Zoom Garante (recomendado) ou Ir direto à loja (Não quero Zoom Garante)

☐ Não quero mais ver isto por 7 dias.

Agora só falta o seu e-mail pra gente te identificar.

Seu email

☑ Quero receber as ofertas do Zoom

☑ Li e aceito o Termo de uso e Política de privacidade


Facebook
Google
O que é o Zoom Garante?
Se a loja não entregar, a gente resolve.
Só quem tem as lojas mais confiáveis pode garantir: se o seu produto não chegar, a gente resolve direto com a loja ou devolve seu dinheiro em até R$3.000.
E é grátis, sem custo nenhum pra você.
Importante: Você precisa estar identificado no Zoom antes de acessar a loja no dia da compra.
Pronto,
Agora você está com o Zoom Garante ativado.
Ir à loja
X Zoom Garante
Oi, .
Agora você já pode contar com o Zoom Garante em sua compra.
OK, ir à loja
Zoom Garante - Se a loja não entregar, a gente resolve.
Aproveite este serviço e compre tranquilo, até em lojas que ainda não conhece. É de graça :)
Ativar agora


Saiba mais
Zoom Garante
top

Anexo 2 – Página Filtro QBM (Consulta: Tablet)

- Informática
  - Tablet

Home › Informática › Tablet › Zoom indica ›

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Zoom indica em Tablet

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Sony S, Samsung Galaxy Tab 10.1, iPad 3 ou Motorola Xoom 2? Saiba qual tablet sai na frente.

Na hora de comprar um tablet, com tantas boas opções disponíveis no mercado, você fica até em dúvida na hora de escolher o modelo ideal para você? Sim, pois modelos não faltam, e muitas vezes ficamos até perdidos na hora de decidir o melhor tablet para nossas tarefas do dia a dia.

Contudo, é sempre bom lembrar que a primeira coisa que você precisa fazer é perceber qual uso você vai dar para o tablet. Ele será só para entrar nas redes sociais? Para ver filmes? Ou mais para jogar? Com essas questões em mente fica mais fácil saber qual será o tablet com a configuração certa para suas atividades.

Para ajudar você, selecionamos alguns tablets de marcas bem conhecidas para uma disputa: o tablet Sony S, o Samsung Galaxy Tab 10.1, o Novo iPad (ou iPad 3) e o Motorola Xoom 2. Confira o que cada modelo oferece de mais interessante e faça a escolha certa.

Depois de conhecer as principais configurações desses tablets, como conexões de internet, câmeras e sistemas operacionais, fica mais prático saber o modelo que entrega os recursos que você tanto precisa. E o melhor de tudo é saber que aqui no Zoom você encontra muitos outros tablets, sendo fácil descobrir o melhor modelo para suas tarefas.

Outros modelos de tablets para você

- Tablet Samsung Galaxy Tab 3 TV 3G 8 GB TFT 7" Android 4.1 (Jelly Bean) 3 MP SM-T211M

  a partir de: R$ 1.406,14 em 2 lojas

  Veja mais
- Tablet Apple iPad Mini 16 GB 7.9" Wi-Fi 5 MP

  a partir de: R$ 991,03 em 4 lojas

  Veja mais

  Ver mais
- Tablet Samsung Galaxy Tab S 16 GB 8,4" Android 4.4...

  Ver mais
- Tablet Samsung Galaxy Tab 3 Lite 3G 8 GB LCD 7"...

  Ver mais

Top 5 em tablet

+ em Tablet

Guia de compras Tá na dúvida? Dê um Zoom em Tablet

- O que posso fazer com um tablet?

  Descubra o que faz um tablet e aproveite todos os seus recursos!
- Por que ter um Tablet?

  Saiba os diferenciais que você vai ter ao comprar um tablet.
- Tablet ou notebook: qual escolher?

  Descubra o ideal para suas tarefas: notebook ou tablet?
- Qual a memória interna ideal para o seu tablet?

  Descubra qual memória interna do tablet não vai deixar você na mão.

Ver todas as matérias

- Tudo sobre os tablets Samsung Galaxy

  Com tantas opções, veja qual tablet Samsung é ideal para o seu dia a dia.
- O que fazer para a bateria do tablet Android durar mais?

  Veja dicas para melhorar a duração da bateria do tablet Android.
- Conheça os tablets com celular

  Um tablet com celular ajuda na comunicação. Confira alguns modelos.

Ver todas as matérias

Guia de compras de outros produtos

Anterior
Próxima
top

Apple e Samsung, por exemplo, que já são marcas conhecidas por entregarem smartphones de sucesso, mantêm o mesmo caminho no mundo dos tablets, oferecendo, geração após geração, modelos cada vez mais modernos.

Então, se você quiser conhecer um tablet Samsung ou um iPad, no Zoom tem vários modelos para você escolher, e até mesmo de outras marcas. Aproveite!

Ah, e se você é daqueles que adora fazer pesquisas antes de comprar qualquer produto, aproveite para comparar preços. Assim você conhece as melhores ofertas e faz a escolha certa! Aliás, não custa lembrar: aqui você encontra durante todo o ano muitas ofertas, e não apenas no dia da Black Friday, data já conhecida por oferecer um dia de grandes descontos na internet.

Veja produtos relacionados em Tablet

- Tablet Samsung Galaxy Tab 3 Lite 8 GB LCD 7" Android...

- Tablet Samsung Galaxy Tab 3 Lite SM-T110 8 GB Android 7"

  19 Avaliações
- Tablet LG G Pad V700 16 GB Android 10,1" 5 MP

  11 Avaliações
- Tablet HP 1201 8 GB Android 7,1" 2 MP 4.2 (Jelly Bean Plus)

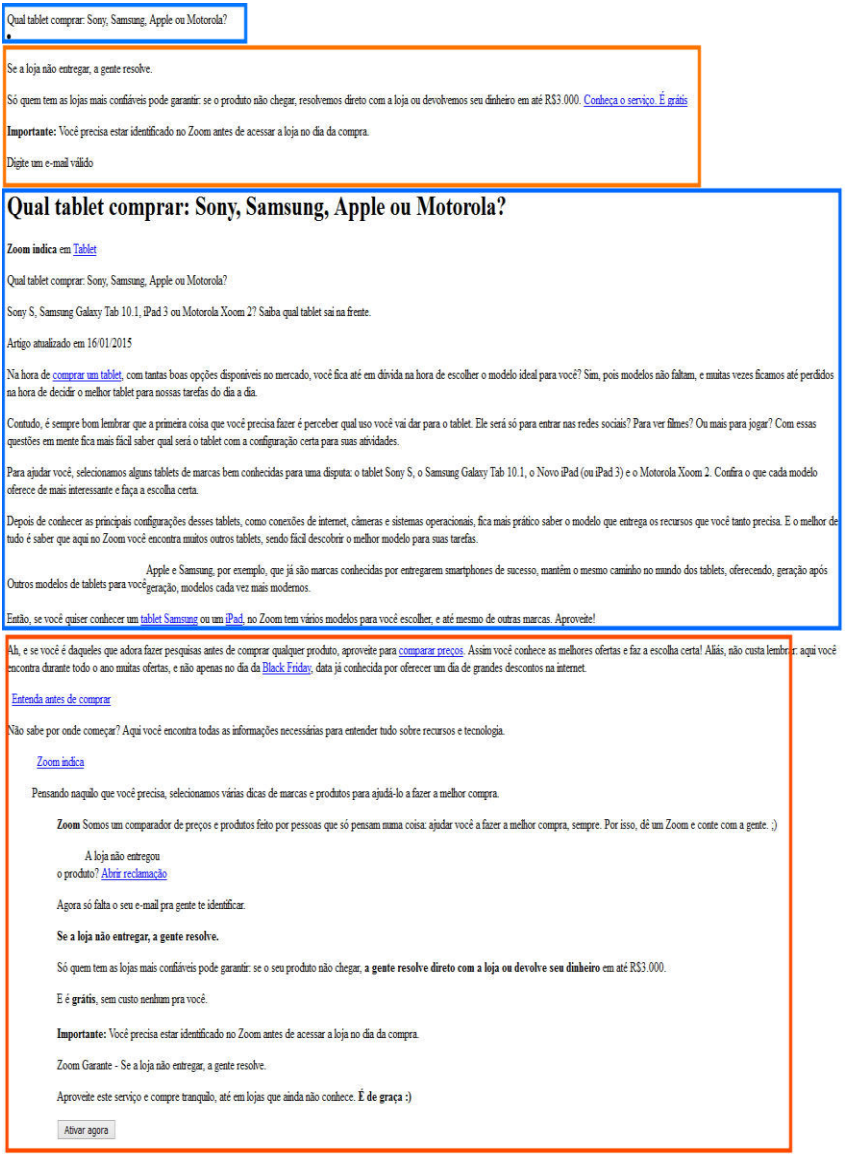
  4 Avaliações
- Tablet Asus Fonepad 7 ME372CG 8 GB 3G Android

  24 Avaliações
- Tablet LG G Pad V400 8 GB Android 7" 3 MP

  12 Avaliações

- Tablet

Anexo 3 – Página Filtro CETR (Consulta: Tablet)



Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Se a loja não entregar, a gente resolve.

Só quem tem as lojas mais confiáveis pode garantir: se o produto não chegar, resolvemos direto com a loja ou devolvemos seu dinheiro em até R$3.000. Conheça o serviço. É grátis

Importante: Você precisa estar identificado no Zoom antes de acessar a loja no dia da compra.

Digite um e-mail válido

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Zoom indica em Tablet

Qual tablet comprar: Sony, Samsung, Apple ou Motorola?

Sony S, Samsung Galaxy Tab 10.1, iPad 3 ou Motorola Xoom 2? Saiba qual tablet sai na frente.

Artigo atualizado em 16/01/2015

Na hora de comprar um tablet, com tantas boas opções disponíveis no mercado, você fica até em dúvida na hora de escolher o modelo ideal para você? Sim, pois modelos não faltam, e muitas vezes ficamos até perdidos na hora de decidir o melhor tablet para nossas tarefas do dia a dia.

Contudo, é sempre bom lembrar que a primeira coisa que você precisa fazer é perceber qual uso você vai dar para o tablet. Ele será só para entrar nas redes sociais? Para ver filmes? Ou mais para jogar? Com essas questões em mente fica mais fácil saber qual será o tablet com a configuração certa para suas atividades.

Para ajudar você, selecionamos alguns tablets de marcas bem conhecidas para uma disputa: o tablet Sony S, o Samsung Galaxy Tab 10.1, o Novo iPad (ou iPad 3) e o Motorola Xoom 2. Confira o que cada modelo oferece de mais interessante e faça a escolha certa.

Depois de conhecer as principais configurações desses tablets, como conexões de internet, câmeras e sistemas operacionais, fica mais prático saber o modelo que entrega os recursos que você tanto precisa. E o melhor de tudo é saber que aqui no Zoom você encontra muitos outros tablets, sendo fácil descobrir o melhor modelo para suas tarefas.

Apple e Samsung, por exemplo, que já são marcas conhecidas por entregarem smartphones de sucesso, mantêm o mesmo caminho no mundo dos tablets, oferecendo, geração após geração, modelos cada vez mais modernos.
Outros modelos de tablets para você

Então, se você quiser conhecer um tablet Samsung ou um iPad, no Zoom tem vários modelos para você escolher, e até mesmo de outras marcas. Aproveite!

Ah, e se você é daqueles que adora fazer pesquisas antes de comprar qualquer produto, aproveite para comparar preços. Assim você conhece as melhores ofertas e faz a escolha certa! Aliás, não custa lembrar: aqui você encontra durante todo o ano muitas ofertas, e não apenas no dia da Black Friday, data já conhecida por oferecer um dia de grandes descontos na internet.

Entenda antes de comprar

Não sabe por onde começar? Aqui você encontra todas as informações necessárias para entender tudo sobre recursos e tecnologia.

Zoom indica

Pensando naquilo que você precisa, selecionamos várias dicas de marcas e produtos para ajudá-lo a fazer a melhor compra.

Zoom Somos um comparador de preços e produtos feito por pessoas que só pensam numa coisa: ajudar você a fazer a melhor compra, sempre. Por isso, dê um Zoom e conte com a gente. ;)

A loja não entregou
o produto? Abrir reclamação

Agora só falta o seu e-mail pra gente te identificar.

Se a loja não entregar, a gente resolve.

Só quem tem as lojas mais confiáveis pode garantir: se o seu produto não chegar, **a gente resolve direto com a loja ou devolve seu dinheiro** em até R$3.000.

E é **grátis**, sem custo nenhum pra você.

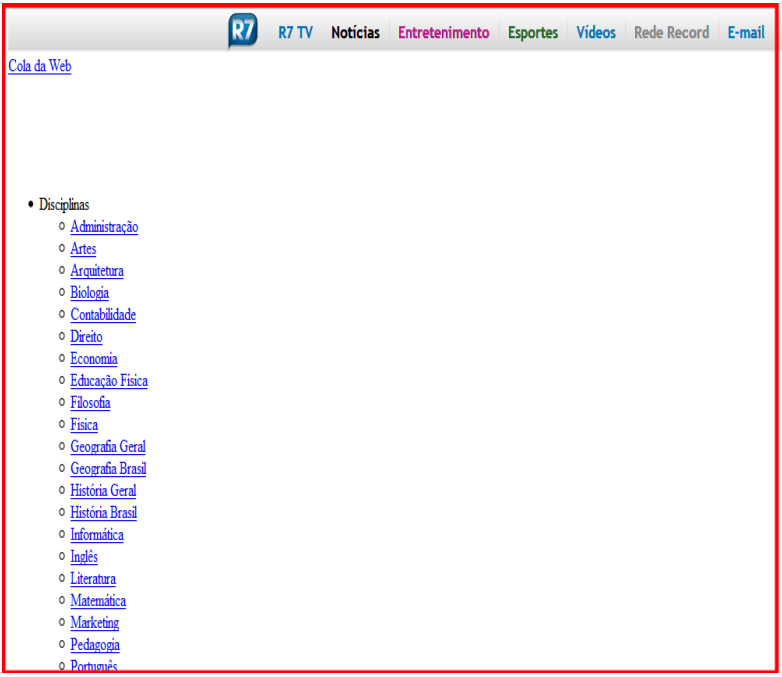
Importante: Você precisa estar identificado no Zoom antes de acessar a loja no dia da compra.

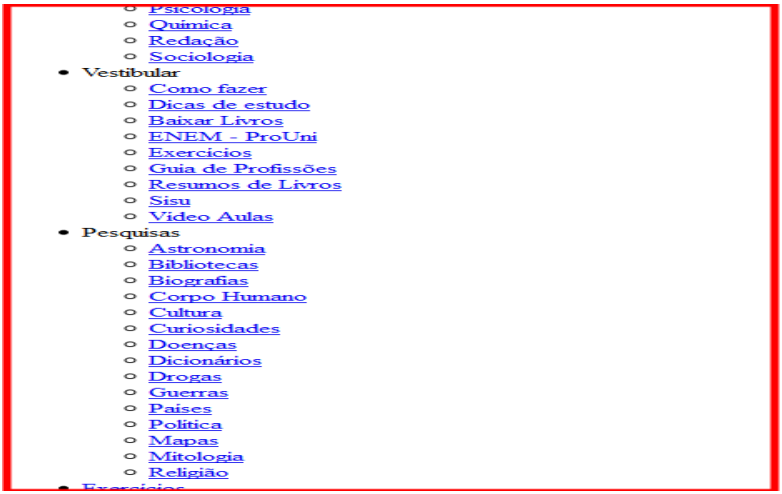
Zoom Garante - Se a loja não entregar, a gente resolve.

Aproveite este serviço e compre tranquilo, até em lojas que ainda não conhece. **É de graça :)**

Ativar agora

Anexo 4 – Página Original (Consulta: Segunda Guerra Mundial)






Biologia
Física
Geografia
História
Inglês
Matemática
Português
Química
Livros
Download de Livros
Resumos de Livros
Guia de Profissões
Google Pesquisa Personalizada
Home » História » Guerras » Segunda Guerra Mundial
Segunda Guerra Mundial
G+1
Tweet
O fato conhecido como Segunda Guerra Mundial, transcorrido entre 1939 e 1945, nada mais foi que uma configuração da guerra anterior.
Sabe-se que ela começou na Europa, mais particularmente devido a uma ofensiva da Alemanha, sendo uma das principais causas o Tratado de Versalhes.
Este tratado assinado em 1919 e que encerrou oficialmente a Primeira Guerra Mundial determinava que a Alemanha assumisse a responsabilidade por ter causado a Primeira Guerra e obrigava o país a pagar uma dívida aos países que saíram prejudicados, além de outras exigências como o impedimento de formar um exército reforçado e o reconhecimento da independência da Áustria.
Estas situações trouxeram revolta aos alemães, que consideraram tais obrigações uma verdadeira humilhação.

A EUROPA ÀS VÉSPERAS DO CONFLITO

Tendo em vista que o fato de que a Segunda Guerra Mundial foi continuação da anterior, deve-se buscar suas raízes mais profundas na forma como a Europa se organizou no fim da Primeira Guerra e após a crise de 1929.

É evidente que os vencidos e vencedores não se entenderam na mesa de negociações. Os vários tratados de paz, sobretudo o de Versalhes, mostraram claramente que a Alemanha foi considerada, pelos vitoriosos, a única culpada pela guerra e, portanto, humilhada com a perda de suas riquezas, de seu território, de seu exército e de sua população.

Segunda Guerra Mundial

Sobre as causas que influenciaram para o início de uma guerra que disseminou milhares de pessoas, sendo considerado um genocídio, foram várias.

O primeiro fator foi o surgimento, na década de 30, de governos totalitários regidos em sua maioria pelo fascismo, como o de Hitler que pretendia expandir o território do país germânico desrespeitando o Tratado de Versalhes.

A Alemanha enfrentando grave crise econômica, juntamente com Itália uniu-se ao Japão e formaram o Eixo que tinha como principal objetivo o de expandir território conquistando territórios vizinhos e ilhas próximas às suas regiões, fazendo um acordo com características militares e com planos em acordo para executar estas conquistas.

O primeiro território invadido no ano de 1939 pela Alemanha foi a Polônia e imediatamente depois os países França e Inglaterra e depois de algum tempo também União Soviética, Inglaterra e Estados Unidos se uniram para declarar guerra à Alemanha formando o grupo dos Aliados.

Estes aliados foram fundamentais para o fim da segunda guerra mundial, do contrário nem saberíamos a que ponto os governantes tiranos teriam chegado.

FASES DA SEGUNDA GUERRA MUNDIAL

As forças em luta nesta guerra ficaram conhecidas como **Forças de Eixo,** resultantes do Pacto de Berlim, Roma e Tóquio, assinado entre Alemanha, Itália e Japão em setembro de 1940; e **Força Aliada,** Inglaterra, França e depois União Soviética, Estados Unidos, Brasil e outros países de menor expressão.

Essa guerra mais que a anterior, atingiu os cinco continentes. Somente o continente americano não serviu diretamente como palco de luta.

Para fins de estudo, pode-se dividir a Segunda Guerra Mundial em duas grandes fases.

A primeira compreendeu os anos iniciais (setembro a 1939 a junho de 1942), quando no continente europeu, as divisões alemãs agiram praticamente como um rolo compressor sobre a Força Aliada. A partir de 1941, as tropas japonesas fizeram o mesmo no Extremo Oriente (Ofensiva do Eixo Vitoriosa).

A segunda fase começou neste período e foi até os anos finais, quando a partir da entrada dos Estados Unidos (Pearl Harbor – 1941) e da (batalha de Stalingrado - 1943) iniciou-se a Contenção da Ofensiva do Eixo, seguida da derrocada final com a vitória dos Aliados.

FATOS HISTÓRICOS IMPORTANTES

• 1939-1941: vitórias do Eixo, lideradas pelas forças armadas da Alemanha – conquistou o Norte da França, Iugoslávia, Polônia, Ucrânia, Noruega e territórios no norte da África. O Japão anexou a Manchúria, enquanto a Itália conquistava a Albânia e territórios da Líbia.

• 1941: Japão ataca a base militar norte-americana Pearl Harbor – EUA entram na Guerra como Aliados.

• 1941-1945: derrotas do Eixo, iniciadas com as perdas sofridas pelos alemães no rigoroso inverno russo – regressão das forças do Eixo que sofrem derrotas seguidas. Com a entrada dos EUA, os aliados ganharam força nas frentes de batalhas.

O Brasil participa diretamente, enviando para a Itália os pracinhas da FEB, Força Expedicionária Brasileira. Os cerca de 25 mil soldados brasileiros conquistam a região, somando uma importante vitória ao lado dos Aliados.

FINAL DA GUERRA E CONSEQUÊNCIAS

Em 1945, Alemanha e Itália se rendem. O Japão, último país a assinar o tratado de rendição, ainda sofreu um forte ataque dos Estados Unidos, que despejou bombas atômicas sobre as cidades de Hiroshima e Nagazaki.

Foram milhões de mortos e feridos, cidades destruídas, indústrias e zonas rurais arrasadas e dívidas incalculáveis.

O racismo esteve presente e deixou uma ferida grave, principalmente na Alemanha, onde os nazistas mandaram para campos de concentração e mataram aproximadamente seis milhões de judeus. Foi criada a ONU (Organização das Nações Unidas), cujo objetivo principal seria a manutenção da paz entre as nações.

O PÓS-GUERRA

A Segunda Guerra Mundial provocou uma alteração significativa de forças no mundo. Seu término marcou também o surgimento de duas superpotências.

De um lado, liderando as democracias liberais, estavam os Estados Unidos, de outro, o bloco socialista, com a URS, exercendo um papel destacado. A partir de 1949, esse bloco foi ampliado com a incorporação da China.

O final da guerra conheceu também a desagregação dos impérios coloniais que ainda sobreviviam, no processo conhecido como descolonização.

Para definir a nova relação de forças internacionais, cunharam-se duas expressões: superpotências e bipolarização – mostrando que o planeta se encontrava dividido em duas zonas de influência econômica, política e ideológica, controladas respectivamente pelos EUA e URSS.

Finalmente, os avanços tecnológicos provocados pela guerra resultaram em numerosas aplicações pacíficas, que vão desde a penicilina até o radar ou a propulsão a jato para os aviões.

REFERÊNCIAS

ARRUDA. José Jobson de A. PILETTI. Nelson. **Toda a História. História Geral e História do Brasil**. Editora Ática: 9ª Edição.

COTRIM. Gilberto. **História Global: Brasil e Geral**. Editora Saraiva. Volume Único. 7ª Edição: 2003. 1ª Tiragem, 2003.

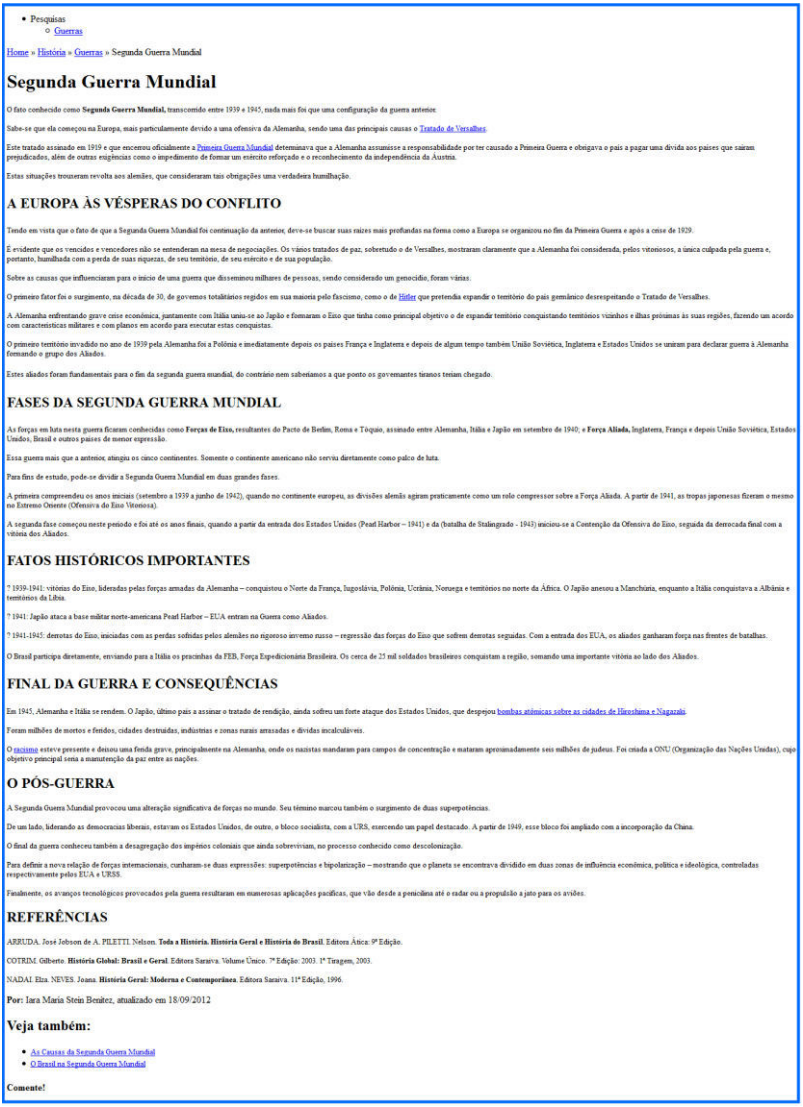
NADAI. Elza. NEVES. Joana. **História Geral: Moderna e Contemporânea**. Editora Saraiva. 11ª Edição, 1996.

Por: Iara Maria Stein Benitez, atualizado em 18/09/2012

Veja também:

- As Causas da Segunda Guerra Mundial
- O Brasil na Segunda Guerra Mundial

Anexo 5 – Página Filtro QBM (Consulta: Segunda Guerra Mundial)



- Pesquisas
  - Guerras

Home » História » Guerras » Segunda Guerra Mundial

Segunda Guerra Mundial

O fato conhecido como **Segunda Guerra Mundial,** transcorrido entre 1939 e 1945, nada mais foi que uma configuração da guerra anterior.

Sabe-se que ela começou na Europa, mais particularmente devido a uma ofensiva da Alemanha, sendo uma das principais causas o Tratado de Versalhes.

Este tratado assinado em 1919 e que encerrou oficialmente a Primeira Guerra Mundial determinava que a Alemanha assumisse a responsabilidade por ter causado a Primeira Guerra e obrigava o país a pagar uma dívida aos países que saíram prejudicados, além de outras exigências como o impedimento de formar um exército reforçado e o reconhecimento da independência da Áustria.

Estas situações trouxeram revolta aos alemães, que consideraram tais obrigações uma verdadeira humilhação.

A EUROPA ÀS VÉSPERAS DO CONFLITO

Tendo em vista que o fato de que a Segunda Guerra Mundial foi continuação da anterior, deve-se buscar suas raízes mais profundas na forma como a Europa se organizou no fim da Primeira Guerra e após a crise de 1929.

É evidente que os vencidos e vencedores não se entenderam na mesa de negociações. Os vários tratados de paz, sobretudo o de Versalhes, mostraram claramente que a Alemanha foi considerada, pelos vitoriosos, a única culpada pela guerra e, portanto, humilhada com a perda de suas riquezas, de seu território, de seu exército e de sua população.

Sobre as causas que influenciaram para o início de uma guerra que disseminou milhares de pessoas, sendo considerado um genocídio, foram várias.

O primeiro fator foi o surgimento, na década de 30, de governos totalitários regidos em sua maioria pelo fascismo, como o de Hitler que pretendia expandir o território do país germânico desrespeitando o Tratado de Versalhes.

A Alemanha enfrentando grave crise econômica, juntamente com Itália unia-se ao Japão e fomaram o Eixo que tinha como principal objetivo o de expandir território conquistando territórios vizinhos e ilhas próximas às suas regiões, fazendo um acordo com características militares e com planos em acordo para executar estas conquistas.

O primeiro território invadido no ano de 1939 pela Alemanha foi a Polônia e imediatamente depois os países França e Inglaterra e depois de algum tempo também União Soviética, Inglaterra e Estados Unidos se uniram para declarar guerra à Alemanha formando o grupo dos Aliados.

Estes aliados foram fundamentais para o fim da segunda guerra mundial, do contrário nem sabemos a que ponto os governantes tiranos teriam chegado.

FASES DA SEGUNDA GUERRA MUNDIAL

As forças em luta nesta guerra ficaram conhecidas como **Forças de Eixo,** resultantes do Pacto de Berlim, Roma e Tóquio, assinado entre Alemanha, Itália e Japão em setembro de 1940; e **Força Aliada,** Inglaterra, França e depois União Soviética, Estados Unidos, Brasil e outros países de menor expressão.

Essa guerra mais que a anterior, atingiu os cinco continentes. Somente o continente americano não serviu diretamente como palco de luta.

Para fins de estudo, pode-se dividir a Segunda Guerra Mundial em duas grandes fases.

A primeira compreendeu os anos iniciais (setembro a 1939 a junho de 1942), quando no continente europeu, as divisões alemãs agiram praticamente como um rolo compressor sobre a Força Aliada. A partir de 1941, as tropas japonesas fizeram o mesmo no Extremo Oriente (Ofensiva do Eixo Vitoriosa).

A segunda fase começou neste período e foi até os anos finais, quando a partir da entrada dos Estados Unidos (Pearl Harbor – 1941) e da (batalha de Stalingrado - 1943) iniciou-se a Contenção da Ofensiva do Eixo, seguida da derrocada final com a vitória dos Aliados.

FATOS HISTÓRICOS IMPORTANTES

? 1939-1941: vitórias do Eixo, lideradas pelas forças armadas da Alemanha – conquistou o Norte da França, Iugoslávia, Polônia, Ucrânia, Noruega e territórios no norte da África. O Japão anexou a Manchúria, enquanto a Itália conquistava a Albânia e territórios da Líbia.

? 1941: Japão ataca a base militar norte-americana Pearl Harbor – EUA entram na Guerra como Aliados.

? 1941-1945: derrotas do Eixo, iniciadas com as perdas sofridas pelos alemães no rigoroso inverno russo – regressão das forças do Eixo que sofrem derrotas seguidas. Com a entrada dos EUA, os aliados ganharam força nas frentes de batalhas.

O Brasil participa diretamente, enviando para a Itália os pracinhas da FEB, Força Expedicionária Brasileira. Os cerca de 25 mil soldados brasileiros conquistam a região, somando uma importante vitória ao lado dos Aliados.

FINAL DA GUERRA E CONSEQUÊNCIAS

Em 1945, Alemanha e Itália se rendem. O Japão, último país a assinar o tratado de rendição, ainda sofreu um forte ataque dos Estados Unidos, que despejou bombas atômicas sobre as cidades de Hiroshima e Nagazaki.

Foram milhões de mortos e feridos, cidades destruídas, indústrias e zonas rurais arrasadas e dívidas incalculáveis.

O racismo esteve presente e deixou uma ferida grave, principalmente na Alemanha, onde os nazistas mandaram para campos de concentração e mataram aproximadamente seis milhões de judeus. Foi criada a ONU (Organização das Nações Unidas), cujo objetivo principal seria a manutenção da paz entre as nações.

O PÓS-GUERRA

A Segunda Guerra Mundial provocou uma alteração significativa de forças no mundo. Seu término marcou também o surgimento de duas superpotências.

De um lado, liderando as democracias liberais, estavam os Estados Unidos, de outro, o bloco socialista, com a URS, exercendo um papel destacado. A partir de 1949, esse bloco foi ampliado com a incorporação da China.

O final da guerra conheceu também a desagregação dos impérios coloniais que ainda sobreviviam, no processo conhecido como descolonização.

Para definir a nova relação de forças internacionais, cunharam-se duas expressões: superpotências e bipolarização – mostrando que o planeta se encontrava dividido em duas zonas de influência econômica, política e ideológica, controladas respectivamente pelos EUA e URSS.

Finalmente, os avanços tecnológicos provocados pela guerra resultaram em numerosas aplicações pacíficas, que vão desde a penicilina até o radar ou a propulsão a jato para os aviões.

REFERÊNCIAS

ARRUDA. José Jobson de A. PILETTI. Nelson. **Toda a História. História Geral e História do Brasil**. Editora Ática. 9ª Edição.

COTRIM. Gilberto. **História Global: Brasil e Geral**. Editora Saraiva. Volume Único. 7ª Edição: 2003. 1ª Tiragem, 2003.

NADAI. Elza. NEVES. Joana. **História Geral: Moderna e Contemporânea**. Editora Saraiva. 11ª Edição, 1996.

Por: Iara Maria Stein Benitez, atualizado em 18/09/2012

Veja também:

- As Causas da Segunda Guerra Mundial
- O Brasil na Segunda Guerra Mundial

Comente!

Segunda Guerra Mundial - Cola da Web

O fato conhecido como **Segunda Guerra Mundial**, transcorrido entre 1939 e 1945, nada mais foi que uma configuração da guerra anterior.

Sabe-se que ela começou na Europa, mais particularmente devido a uma ofensiva da Alemanha, sendo uma das principais causas o Tratado de Versalhes.

Este tratado assinado em 1919 e que encerrou oficialmente a Primeira Guerra Mundial determinava que a Alemanha assumisse a responsabilidade por ter causado a Primeira Guerra e obrigava o país a pagar uma dívida aos países que saíram prejudicados, além de outras exigências como o impedimento de formar um exército reforçado e o reconhecimento da independência da Áustria.

Estas situações trouxeram revolta aos alemães, que consideraram tais obrigações uma verdadeira humilhação.

A EUROPA ÀS VÉSPERAS DO CONFLITO

Tendo em vista que o fato de que a Segunda Guerra Mundial foi continuação da anterior, deve-se buscar suas raízes mais profundas na forma como a Europa se organizou no fim da Primeira Guerra e após a crise de 1929.

É evidente que os vencidos e vencedores não se entenderam na mesa de negociações. Os vários tratados de paz, sobretudo o de Versalhes, mostraram claramente que a Alemanha foi considerada, pelos vitoriosos, a única culpada pela guerra e, portanto, humilhada com a perda de suas riquezas, de seu território, de seu exército e de sua população.

Segunda Guerra Mundial

Sobre as causas que influenciaram para o início de uma guerra que disseminou milhares de pessoas, sendo considerado um genocídio, foram várias.

O primeiro fator foi o surgimento, na década de 30, de governos totalitários regidos em sua maioria pelo fascismo, como o de Hitler que pretendia expandir o território do país germânico desrespeitando o Tratado de Versalhes.

A Alemanha enfrentando grave crise econômica, juntamente com Itália uniu-se ao Japão e formaram o Eixo que tinha como principal objetivo o de expandir território conquistando territórios vizinhos e ilhas próximas às suas regiões, fazendo um acordo com características militares e com planos em acordo para executar estas conquistas.

O primeiro território invadido no ano de 1939 pela Alemanha foi a Polônia e imediatamente depois os países França e Inglaterra e depois de algun tempo também União Soviética, Inglaterra e Estados Unidos se uniram para declarar guerra à Alemanha formando o grupo dos Aliados.

Estes aliados foram fundamentais para o fim da segunda guerra mundial, do contrário nem saberíamos a que ponto os governantes tiranos teriam chegado.

FASES DA SEGUNDA GUERRA MUNDIAL

As forças em luta nesta guerra ficaram conhecidas como **Forças do Eixo,** resultantes do Pacto de Berlim, Roma e Tóquio, assinado entre Alemanha, Itália e Japão em setembro de 1940; e **Força Aliada,** Inglaterra, França e depois União Soviética, Estados Unidos, Brasil e outros países de menor expressão.

Essa guerra mais que a anterior, atingiu os cinco continentes. Somente o continente americano não serviu diretamente como palco de luta.

Para fins de estudo, pode-se dividir a Segunda Guerra Mundial em duas grandes fases.

A primeira compreendeu os anos iniciais (setembro a 1939 a junho de 1942), quando no continente europeu, as divisões alemãs agiram praticamente como um rolo compressor sobre a Força Aliada. A partir de 1941, as tropas japonesas fizeram o mesmo no Extremo Oriente (Ofensiva do Eixo Vitoriosa).

A segunda fase começou neste período e foi até os anos finais, quando a partir da entrada dos Estados Unidos (Pearl Harbor – 1941) e da (batalha de Stalingrado - 1943) iniciou-se a Contenção da Ofensiva do Eixo, seguida da derrocada final com a vitória dos Aliados.

FATOS HISTÓRICOS IMPORTANTES

- 1939-1941: vitórias do Eixo, lideradas pelas forças armadas da Alemanha – conquistou o Norte da França, Iugoslávia, Polônia, Ucrânia, Noruega e territórios no norte da África. O Japão anexou a Manchúria, enquanto a Itália conquistava a Albânia e territórios da Líbia.
- 1941: Japão ataca a base militar norte-americana Pearl Harbor – EUA entram na Guerra como Aliados.
- 1941-1943: derrotas do Eixo, iniciadas com as perdas sofridas pelos alemães no rigoroso inverno russo – regressão das forças do Eixo que sofrem derrotas seguidas. Com a entrada dos EUA, os aliados ganharam força nas frentes de batalhas.

O Brasil participa diretamente, enviando para a Itália os pracinhas da FEB, Força Expedicionária Brasileira. Os cerca de 25 mil soldados brasileiros conquistam a região, somando uma importante vitória ao lado dos Aliados.

FINAL DA GUERRA E CONSEQUÊNCIAS

Em 1945, Alemanha e Itália se rendem. O Japão, último país a assinar o tratado de rendição, ainda sofreu um forte ataque dos Estados Unidos, que despejou bombas atômicas sobre as cidades de Hiroshima e Nagazaki.

Foram milhões de mortos e feridos, cidades destruídas, indústrias e zonas rurais arrasadas e dívidas incalculáveis.

O racismo esteve presente e deixou uma ferida grave, principalmente na Alemanha, onde os nazistas mandaram para campos de concentração e mataram aproximadamente seis milhões de judeus. Foi criada a ONU (Organização das Nações Unidas), cujo objetivo principal seria a manutenção da paz entre as nações.

O PÓS-GUERRA

A Segunda Guerra Mundial provocou uma alteração significativa de forças no mundo. Seu término marcou também o surgimento de duas superpotências.

De um lado, liderando as democracias liberais, estavam os Estados Unidos, de outro, o bloco socialista, com a URS, exercendo um papel destacado. A partir de 1949, esse bloco foi ampliado com a incorporação da China.

O final da guerra conheceu também a desagregação dos impérios coloniais que ainda sobreviviam, no processo conhecido como descolonização.

Para definir a nova relação de forças internacionais, cunharam-se duas expressões: superpotências e bipolarização – mostrando que o planeta se encontrava dividido em duas zonas de influência econômica, política e ideológica, controladas respectivamente pelos EUA e URSS.

Finalmente, os avanços tecnológicos provocados pela guerra resultaram em numerosas aplicações pacíficas, que vão desde a penicilina até o radar ou a propulsão a jato para os aviões.

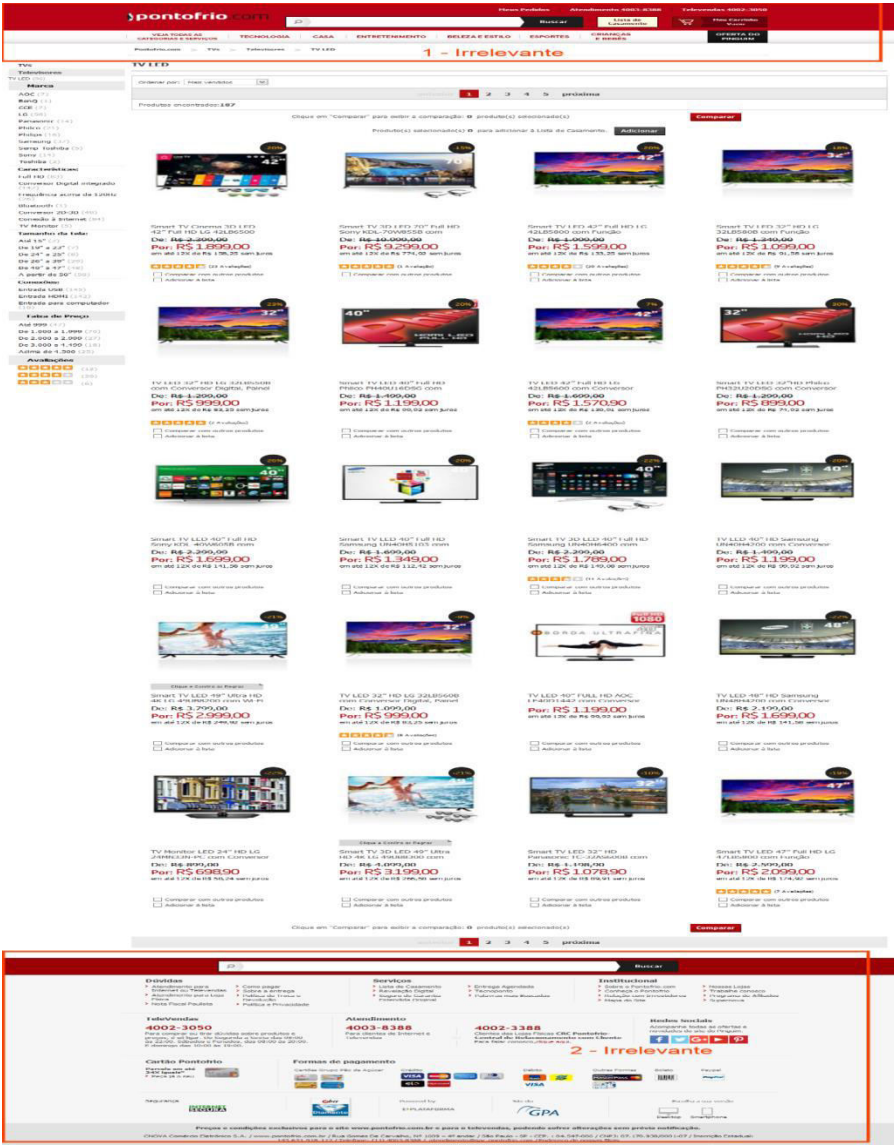
REFERÊNCIAS

ARRUDA, José Jobson de A. PILETTI, Nelson. **Toda a História. História Geral e História do Brasil**. Editora Ática. 9ª Edição.

COTRIM, Gilberto. **História Global: Brasil e Geral**. Editora Saraiva. Volume Único. 7ª Edição: 2003. 1ª Tiragem, 2003.

NADAI, Elza. NEVES, Joana. **História Geral: Moderna e Contemporânea**. Editora Saraiva. 11ª Edição, 1996.

Anexo 7 – Página Original (Consulta: TV Led)



Anexo 8 – Página Filtro QBM (Consulta: TV Led)

- TVs

Buscar
Todo o site
1 - Irrelevante

Pontofrio.com > TVs > Televisores > **TV LED**

TV LED

- **Smart TV Cinema 3D LED 42" Full HD LG 42LB6500 com Sistema WebOS, Painel IPS, Entradas HDMI e USB, 4 Óculos 3D e Controle Smart Magic** Avaliação dos usuários: 45(23 Avaliações)De: **R$ 2.399,00**Por: **R$ 1.899,00**em até 12X de R$ 158,25 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV 3D LED 70" Full HD Sony KDL-70W855B com Motionflow 480hz, Processador X-Reality Pro, Wi-Fi e 2 Óculos 3D** Avaliação dos usuários: 50(1 Avaliação)De: **R$ 10.999,00**Por: **R$ 9.299,00**em até 12X de R$ 774,92 sem juros Comprar
  - Comparar com outros produtos
- **Smart TV LED 42" Full HD LG 42LB5800 com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI** Avaliação dos usuários: 40(20 Avaliações)De: **R$ 1.999,00**Por: **R$ 1.599,00**em até 12X de R$ 133,25 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 32" HD LG 32LB580B com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI** Avaliação dos usuários: 45(9 Avaliações)De: **R$ 1.349,00**Por: **R$ 1.099,00**em até 12X de R$ 91,58 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 32" HD LG 32LB550B com Conversor Digital, Painel IPS, Entradas HDMI e USB** Avaliação dos usuários: 50(2 Avaliações)De: **R$ 1.299,00**Por: **R$ 999,00**em até 12X de R$ 83,25 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 40" Full HD Philco PH40U16DSG com Conversor Digital, Tecnologia Ginga e Entradas HDMI e USB** De: **R$ 1.499,00**Por: **R$ 1.199,00**em até 12X de R$ 99,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 42" Full HD LG 42LB5600 com Conversor Digital, Painel IPS, Entradas HDMI e USB** Avaliação dos usuários: 40(2 Avaliações)De: **R$ 1.699,00**Por: **R$ 1.570,90**em até 12X de R$ 130,91 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 32"HD Philco PH32U20DSG com Conversor Digital, Tecnologia Ginga e Entradas HDMI e USB** De: **R$ 1.299,00**Por: **R$ 899,00**em até 12X de R$ 74,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 40" Full HD Sony KDL-40W605B com Motionflow 240hz, Processador X-Reality PRO, Wi-Fi e Entradas HDMI e USB** De: **R$ 2.299,99**Por: **R$ 1.699,00**em até 12X de R$ 141,58 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 40" Full HD Samsung UN40H5103 com Função Futebol, ConnectShare Movie, Entradas HDMI e USB e Wi-Fi** De: **R$ 1.699,00**Por: **R$ 1.349,00**em até 12X de R$ 112,42 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV 3D LED 40" Full HD Samsung UN40H6400 com Função Futebol, 480Hz Clear Motion Rate, Wi-Fi e 2 Óculos 3D** Avaliação dos usuários: 35(11 Avaliações)De: **R$ 2.299,00**Por: **R$ 1.789,00**em até 12X de R$ 149,08 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 40" HD Samsung UN40H4200 com Conversor Digital, Função Futebol, ConnectShare Movie e Entradas USB e HDMI** De: **R$ 1.499,00**Por: **R$ 1.199,00**em até 12X de R$ 99,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 49" Ultra HD 4K LG 49UB8200 com Wi-Fi Integrado, Time Machine II, Painel Futebol e Controle Remoto Smart Magic** De: **R$ 3.799,00**Por: **R$ 2.999,00**em até 12X de R$ 249,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 32" HD LG 32LB560B com Conversor Digital, Painel IPS, Entradas HDMI e USB** Avaliação dos usuários: 45(8 Avaliações)De: **R$ 1.099,00**Por: **R$ 999,00**em até 12X de R$ 83,25 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 40" FULL HD AOC LE40D1442 com Conversor Digital, Entradas HDMI e USB** Por: **R$ 1.199,00**em até 12X de R$ 99,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV LED 48" HD Samsung UN48H4200 com Conversor Digital, Função Futebol, ConnectShare Movie e Entradas USB e HDMI** De: **R$ 2.199,00**Por: **R$ 1.699,00**em até 12X de R$ 141,58 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **TV Monitor LED 24" HD LG 24MN33N-PC com Conversor Digital, Time Machine Ready, Entradas HDMI e USB** De: **R$ 899,00**Por: **R$ 698,90**em até 12X de R$ 58,24 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV 3D LED 49" Ultra HD 4K LG 49UB8300 com Wi-Fi Integrado, Time Machine II, Painel Futebol, 4 Óculos Cinema 3D e Controle Smart Magic** De: **R$ 4.099,00**Por: **R$ 3.199,00**em até 12X de R$ 266,58 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 32" HD Panasonic TC-32AS600B com Conversor Digital, Wi-Fi, Entradas HDMI e USB** De: **R$ 1.198,90**Por: **R$ 1.078,90**em até 12X de R$ 89,91 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista
- **Smart TV LED 47" Full HD LG 47LB5800 com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI** Avaliação dos usuários: 50(7 Avaliações)De: **R$ 2.599,00**Por: **R$ 2.099,00**em até 12X de R$ 174,92 sem juros Comprar
  - Comparar com outros produtos
  - Adicionar à lista

TVs

Televisores

- TV LED (96)

Características:

- TV Monitor (5)

Tamanho da tela:

Conexões:

Buscar
Todo o site
2 - Irrelevante

Anexo 9 – Página Filtro CETR (Consulta: TV Led)

Smart TV Cinema 3D LED 42 Full HD LG 42LB6500 com Sistema WebOS, Painel IPS, Entradas HDMI e USB, 4 Óculos 3D e Controle Smart Magic Avaliação dos usuários: 45(23 Avaliações)De: R$ 2.399,00Por: R$ 1.899,00em até 12X de R$ 158,25 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV 3D LED 70 Full HD Sony KDL-70W855B com Motionflow 480hz, Processador X-Reality Pro, Wi-Fi e 2 Óculos 3D Avaliação dos usuários: 50(1 Avaliação)De: R$ 10.999,00Por: R$ 9.299,00em até 12X de R$ 774,92 sem juros Comprar

Comparar com outros produtos

Smart TV LED 42 Full HD LG 42LB5800 com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI Avaliação dos usuários: 40(20 Avaliações)De: R$ 1.999,00Por: R$ 1.599,00em até 12X de R$ 133,25 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 32 HD LG 32LB580B com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI Avaliação dos usuários: 45(9 Avaliações)De: R$ 1.349,00Por: R$ 1.099,00em até 12X de R$ 91,58 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

TV LED 32 HD LG 32LB550B com Conversor Digital, Painel IPS, Entradas HDMI e USB Avaliação dos usuários: 50(2 Avaliações)De: R$ 1.299,00Por: R$ 999,00em até 12X de R$ 83,25 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 40 Full HD Philco PH40U16DSG com Conversor Digital, Tecnologia Ginga e Entradas HDMI e USB De: R$ 1.499,00Por: R$ 1.199,00em até 12X de R$ 99,92 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

TV LED 42 Full HD LG 42LB5600 com Conversor Digital, Painel IPS, Entradas HDMI e USB Avaliação dos usuários: 40(2 Avaliações)De: R$ 1.699,00Por: R$ 1.570,90em até 12X de R$ 130,91 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 32 HD Philco PH32U20DSG com Conversor Digital, Tecnologia Ginga e Entradas HDMI e USB De: R$ 1.299,00Por: R$ 899,00em até 12X de R$ 74,92 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 40 Full HD Sony KDL-40W605B com Motionflow 240hz, Processador X-Reality PRO, Wi-Fi e Entradas HDMI e USB De: R$ 2.299,99Por: R$ 1.699,00em até 12X de R$ 141,58 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 40 Full HD Samsung UN40H5103 com Função Futebol, ConnectShare Movie, Entradas HDMI e USB e Wi-Fi De: R$ 1.699,00Por: R$ 1.349,00em até 12X de R$ 112,42 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV 3D LED 40◆ Full HD Samsung UN40H6400 com Função Futebol, 480Hz Clear Motion Rate, Wi-Fi e 2 Óculos 3D Avaliação dos usuários: 35(11 Avaliações)De: R$ 2.299,00Por: R$ 1.789,00em até 12X de R$ 149,08 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

TV LED 40◆ HD Samsung UN40H4200 com Conversor Digital, Função Futebol, ConnectShare Movie e Entradas USB e HDMI De: R$ 1.499,00Por: R$ 1.199,00em até 12X de R$ 99,92 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

REGULAMENTO DA PROMOÇÃO

Smart TV LED 49◆ Ultra HD 4K LG 49UB8200 com Wi-Fi Integrado, Time Machine II, Painel Futebol e Controle Remoto Smart Magic De: R$ 3.799,00Por: R$ 2.999,00em até 12X de R$ 249,92 sem juros

Comparar com outros produtos

Adicionar à lista

TV LED 32◆ HD LG 32LB560B com Conversor Digital, Painel IPS, Entradas HDMI e USB Avaliação dos usuários: 45(8 Avaliações)De: R$ 1.099,00Por: R$ 999,00em até 12X de R$ 83,25 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

TV LED 40◆ FULL HD AOC LE40D1442 com Conversor Digital, Entradas HDMI e USB Por: R$ 1.199,00em até 12X de R$ 99,92 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

TV Monitor LED 24" HD LG 24MN33N-PC com Conversor Digital, Time Machine Ready, Entradas HDMI e USB De: R$ 899,00Por: R$ 698,90em até 12X de R$ 58,24 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

REGULAMENTO DA PROMOÇÃO

Smart TV 3D LED 49" Ultra HD 4K LG 49UB8300 com Wi-Fi Integrado, Time Machine II, Painel Futebol, 4 Óculos Cinema 3D e Controle Smart Magic De: R$ 4.099,00Por: R$ 3.199,00em até 12X de R$ 266,58 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 32" HD Panasonic TC-32AS600B com Conversor Digital, Wi-Fi, Entradas HDMI e USB De: R$ 1.198,90Por: R$ 1.078,90em até 12X de R$ 89,91 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Smart TV LED 47" Full HD LG 47LB5800 com Função Torcida, Conversor Digital, Wi-Fi, Entradas USB e HDMI Avaliação dos usuários: 50(7 Avaliações)De: R$ 2.599,00Por: R$ 2.099,00em até 12X de R$ 174,92 sem juros Comprar

Comparar com outros produtos

Adicionar à lista

Comparar

Clique em "Comparar" para exibir a comparação: produto(s) selecionado(s)

Televisores

- TV LED (96)
- Philco (21)
- Philips (16)
- Samsung (37)
- Semp Toshiba (5)

Características:

- Full HD (83)
- Conversor Digital integrado (142)
- Frequência acima de 120Hz (26)
- Bluetooth (1)
- Conversor 2D-3D (48)
- Conexão à Internet (84)
- TV Monitor (5)

Tamanho da tela:

- Até 15" (2)
- De 19" a 23" (7)
- De 24" a 25" (8)
- De 26" a 39" (29)
- De 40" a 47" (48)
- A partir de 50" (50)

Conexões:

- Entrada USB (145)
- Entrada HDMI (142)
- Entrada para computador (18)

Faixa de Preço

- Até 999 (47)
- De 1.000 a 1.999 (70)
- De 2.000 a 2.999 (27)
- De 3.000 a 4.499 (18)
- Acima de 4.500 (25)

Todo o site

Dúvidas

- Atendimento para Internet ou Televendas
- Atendimento para Loja Física
- Nota Fiscal Paulista
- Como pagar
- Sobre a entrega
- Política de Troca e Devolução
- Política e Privacidade

Serviços

- Lista de Casamento
- Revelação Digital
- Seguro de Garantia Estendida Original
- Entrega Agendada
- Tecnoponto
- Palavras mais Buscadas

Institucional

- Sobre o Pontofrio.com
- Conheça o Pontofrio
- Relação com Investidores
- Mapa do Site
- Nossas Lojas
- Trabalhe conosco
- Programa de Afiliados
- Supernova

- TeleVendas**4002-3050**

Para comprar ou tirar dúvidas sobre produtos e preços, é só ligar. De Segunda a Sexta das 08:00 às 22:00. Sábados e Feriados, das 08:00 às 20:00. E domingo das 10:00 às 19:00.

- Atendimento**4003-8388**

  Para clientes de Internet e Televendas

  4002-3388

  Clientes das Lojas Físicas **CRC Pontofrio**- Central de Relacionamento com Cliente

  Para falar conosco,clique aqui

- Redes Sociais

  Acompanhe todas as ofertas e novidades do site do Pinguim.

  - Facebook
  - Twitter
  - Encontre-nos no Google+
  - YouTube
  - Pinterest

  Cartão Pontofrio

  Parcele em até
  24X iguais*
  Peça já o seu

  Formas de pagamento

  - Cartões Grupo Pão de Açúcar

    Cartões do Grupo Pão de Açúcar (Extra, Sendas, Compre Bem, Pão de Açúcar)

  - Crédito

    Cartões de crédito Ponto Frio, Visa, Mastercard, American Express, Diners e Aura

  - Débito

    Cartões de débito Itaú, Bradesco e Unibanco; BB Crediário; Visa Electron

  - Outras Formas

    Multicheque e Multcash

  - Boleto

    Boleto bancário

  - Paypal

    Paypal

    - Ebit loja Diamante
    - Powered by

      e-Plataforma

    - Site do

      Grupo Pão de Açúcar

    - Escolha a sua versão

      - **Desktop**

Preços e condições exclusivos para o site www.pontofrio.com.br e para o televendas, podendo sofrer alterações sem prévia notificação.

CNOVA Comércio Eletrônico S.A. / www.pontofrio.com.br / Rua Gomes De Carvalho, Nº 1609 ♦ 4º andar / São Paulo - SP - CEP. : 04.547-006 / CNPJ: 07.170.938/0001-07 / Inscrição Estadual: 143.631.918.112 / Telefone: (11) 4003-8388 / atendimento@sac.pontofrio.com /Endereço de nossas filiais

- **Telefones e Celulares**
  - Celulares Desbloqueados
  - Celulares de Operadoras
  - Smartphones
  - Smartwatches
  - Telefonia Fixa
  - Acessórios para Celulares

  Veja mais
- **Cine e Foto**
  - Câmeras Compactas
  - Câmeras Semi-Profissionais
  - Câmeras Reflex / SLR
  - Filmadoras
  - Pilhas/Baterias e Carregadores
  - Lentes e Filtros
  - Tripés
  - Bolsa e Estojos
  - Adegas
  - Aquecedores
  - Fogões
  - Forno de Micro-ondas
  - Lavadoras
  - Lava e Seca
  - Linha Industrial
  - Refrigeradores / Geladeiras
  - Secadoras

Sua Cozinha de Embutir

- Coifas
- Cooktops
- Forno de Embutir
- **Ar e ventilação**
  - Ar Condicionado
  - Ventiladores e Circuladores
  - Climatizadores
  - Aquecedores
  - Umidificadores de Ar

  Veja mais
- **Móveis**
  - Dormitórios / Quartos
  - Cozinha
  - Sala de Estar
  - Colchões
  - Móveis de Escritório

  Veja mais
- **Eletroportáteis**
  - Aspirador de Pó
  - Bebedouro
  - Cafeteiras
  - Máquinas de Costura
  - Processadores de Alimento

  Veja mais
- **Utilidades Domésticas**
  - Aparelhos de Jantar
  - Baixelas e Travessas
  - Faqueiros
  - Conjunto de Panelas
  - Decoração

  Veja mais
- **Cama, Mesa e Banho**
  - Jogo de Cama
  - Edredom
  - Cobertor e Manta
  - Colcha e Cobre-leito
  - Travesseiro

  Veja mais

- **Ferramentas**
  - Elétricas
  - Manuais
  - Pneumáticas
  - Lavadoras de Pressão
  - Jardinagem

  Veja mais
  - Televisores
  - Dvd & Blu-Ray Players
  - Home Theater
  - Áudio
  - PS Vita
  - PSP
  - Nintendo Wii U
  - Nintendo Wii
  - Nintendo 3DS
  - Nintendo DS
  - Acessórios
  - Download de Jogos
  - Jogos para PC

  Veja mais
- **FILMES E MÚSICAS**
  - Pré-Venda
  - Lançamento
  - Mais Vendidos

  Filmes
  - Minisséries e Séries de TV
  - Ação e Aventura
  - Infantil
  - Drama

  Música
  - Pop e Rock Internacional
  - MPB
  - Sertanejo
  - **Livros**
    - Pré-Venda
    - Lançamento
    - Mais Vendidos
    - Auto-Ajuda e Relacionamentos
    - Biografias
    - Literatura Estrangeira
    - Literatura Nacional
    - Literatura Infantojuvenil
    - Informática e Certificação
    - Didáticos
    - Religião
    - Viagens e Turismo

    Veja mais
  - **Papelaria**
    - Calculadora
    - Mochilas, Estojos & Lancheiras
    - Apresentação / Equipamentos
    - Cofre
    - Escolar / Escritório
    - Papéis, Pastas & Arquivos
    - Escrita & Corretivos
    - Artes & Pinturas
    - Envelopes, Etiquetas, Formulários
    - Festas e Presentes

    Veja mais

- Veja mais
- Televisores
- Dvd & Blu-Ray Players
- Home Theater
- Áudio
  - Barbeadores
  - Pranchas (Chapinhas)
  - Secadores de Cabelo
  - Depiladores
  - Balanças
  - Massageadores
  - Medidores de Pressão
- Veja mais
- **Malas & Mochilas**
  - Malas Avulsas
  - Conjuntos
  - Frasqueiras
  - Sacolas de Viagem
  - Mochilas
  - Bolsas
  - Acessórios
- Veja mais
- **Relógios**
  - Masculinos
  - Unissex
    - Cabelos
    - Corpo e Banho
    - Maquiagem
    - Fitness
    - Bicicletas
    - Camping
    - Monitores Cardíacos
    - Praia e Piscina
    - Vestuário
    - Óculos de Sol
    - Jogos de Mesa
    - Mochilas Esportivas
    - Suplementos Alimentares
  - Veja mais
  -
  - Patins
  - Surf
  - Ténis de Mesa
  - Jogos de Mesa
  - Boxe
  - Jiu-Jitsu
  - Bonecas
  - Bonecos
  - Controle Remoto
  - Esporte infantil
  - Mini Veículos
  - Puericultura
  - Playground
  - Praia e Piscina
  - Tocas e Barracas
- Veja mais
- **Bebês**
  - Carrinhos
  - Bebês Conforto
  - Andadores
  - Berços e Cercados
  - Cadeiras para Automóveis
  - Fraldas
  - Banheiras
  - Brinquedos para Bebê
  - Cadeirões
  - Segurança do Bebê
- Veja mais

- **Volta às aulas universitário**
  - **Volta às aulas universitário**
  - **Cada dia uma oferta especial para você ;)**