

Técnicas de Busca na Web



**Instituto Agronômico
de Campinas**

10 de maio de 2004

Rubens Queiroz de Almeida

queiroz@unicamp.br

A Internet e a Informação



Vantagens

- ❑ Atualizada
- ❑ Acesso fácil
- ❑ Funciona 24 horas ao dia
- ❑ Acesso a uma grande variedade de recursos

Desvantagens

- ❑ Qualidade?
- ❑ Falta de organização
- ❑ Mecanismos de busca imperfeitos?



O que tem na Web

- Livros Online
- Jornais eletrônicos
- Bancos de dados bibliográficos
- Material de texto completo



- Páginas pessoais
- Grupos de discussão e listas
- Audio, video, imagens
- Recursos interativos (blogs, ...)
- Serviços gratuitos e por assinatura

O buscador



- Não consulta muitas páginas
- Apenas 15% consultam mais que uma página
- A maioria lê apenas superficialmente
- Um estudo demonstrou que 10% não sabem como fazer scroll em uma página

Índices

- Google <http://www.google.com>
- Alta Vista <http://www.altavista.com>
- Excite <http://www.excite.com>
- HotBot <http://www.hotbot.lycos.com>

Catálogos

- Listagens criadas por seres humanos
- Visão mais organizada da web
- Limitada
- Maior qualidade
- Exemplos:
 - Yahoo <http://www.yahoo.com>
 - Dmoz <http://dmoz.org>
 - LookSmart <http://www.looksmart.com>

Catálogos

- Coleção de links privados
 - Dicas-L
 - <http://www.Dicas-L.unicamp.br/hotlinks>
 - Palavras chave
 - “websearch bookmarks”

Google

- O melhor indexador da Web
- Ético, simples e objetivo
- Cada vez mais imitado
- Lucrativo
 - Anúncios direcionados
 - Ótimos resultados
- Oferta pública de ações (IPO) em 2004

Google e Inovação

- Inovação constante
 - <http://labs.google.com>
 - Images <http://images.google.com>
 - Froogle <http://froogle.google.com>
 - News <http://news.google.com>
 - Google Sets <http://labs.google.com/sets>
 - Voice Search <http://labs1.google.com/gvs.html>
 - WebQuotes <http://labs.google.com/cgi-bin/webquotes>
 - ...

Busca com Google

- Busca básica e avançada
- Recursos poderosos para refinamento de buscas
- Documentação em português
 - <http://www.google.com.br/intl/pt-BR/help.html>

medicine

Google™

[Web](#) [Imagens](#) [Grupos](#) [Diretório](#)

medicine

Pesquisar

[Pesquisa avançada](#)

[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web

Resultados 1 - 10 de aproximadamente **42,000,000** para **medicine** (0.17 segundos)

Pediatric medicine



Web [Imagens](#) [Grupos](#) [Diretório](#)

pediatric medicine

Pesquisar

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente **1,540,000** para **pediatric medicine** (0.20 segundos)

“pediatric medicine”

Google™

[Web](#) [Imagens](#) [Grupos](#) [Diretório](#)

"pediatric medicine"

Pesquisar

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 21,000 para "pediatric medicine" (0.07 segundos)

"pediatric medicine" +nose



Web [Imagens](#) [Grupos](#) [Diretório](#)

"pediatric medicine" +nose

Pesquisar

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente **1,750** para "pediatric medicine" +nose (0.14 segundos)

"pediatric medicine" +nose -brazil



[Web](#) [Imagens](#) [Grupos](#) [Diretório](#)

"pediatric medicine" +nose -brazil

Pesquisar

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 1,710 para "pediatric medicine" +nose -brazil (0.11

"pediatric medicine" +nose +brazil



Web [Imagens](#) [Grupos](#) [Diretório](#)

"pediatric medicine" +nose +brazil

Pesquisar

[Pesquisa avançada](#)

[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados 1 - 10 de aproximadamente 31 para "pediatric medicine" +nose +brazil (0.12 se

intitle:"busca na web" filetype:ppt



[Web](#) [Imagens](#) [Grupos](#) [Diretório](#)

intitle:"busca na web" filetype:ppt

Pesquisar

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web

Resultados **1 - 1** de **1** para **intitle:"busca na web" filetype:ppt** (0.13 segundos)

Dica: Tente remover as aspas de sua pesquisa para obter mais resultados.

[PPT] [A BIBLIOTECA VIRTUAL COMO MECANISMO DE **BUSCA NA WEB**: O LOCALIZADOR ...](#)

Formato do arquivo: Microsoft Powerpoint 97 - [Ver em HTML](#)
BIBLIOTECA/CIR – FSP/USP - Brasil. A Biblioteca Virtual como Mecanismo de **Busca na Web**: o Localizador de Informação em Saúde Pública na Internet (LIS-SP). ...
www.usp.br/sibi/boletim_inter/download/CRICS6%20Angela.ppt - [Páginas Semelhantes](#)

Links Patrocinados

[Documentos Corporativos](#)

Localize, Compartilhe e Controle: Documentos, Versões, Leitores, etc.
www.e-docs.com.br

[Veja a sua mensagem aqui...](#)

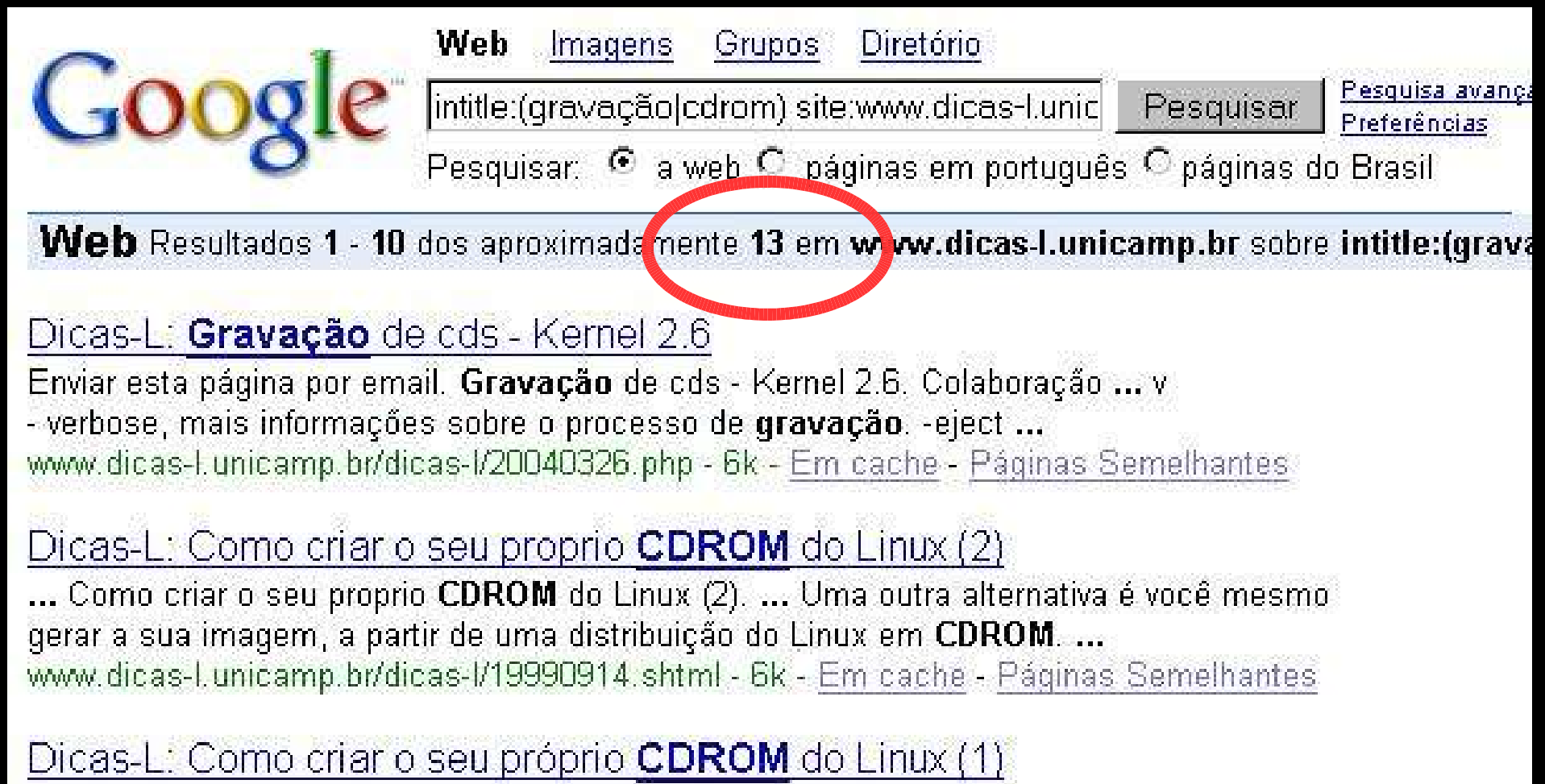
intitle:"busca na web" filetype:pdf

The image shows a Google search interface. At the top left is the Google logo. To its right are navigation links: [Web](#), [Imagens](#), [Grupos](#), and [Diretório](#). Below these is a search input field containing the text "intitle:busca na web" filetype:pdf. To the right of the input field is a "Pesquisar" button. Further right are links for "Pesquisa avançada" and "Preferências". Below the search bar, there are radio buttons for "a web" (selected), "páginas em português", and "páginas do Brasil".

The search results are displayed under the heading "Web". The first result is circled in red and shows "Resultados 1 - 3 de 3 para intitle:busca na web filetype:pdf (0.14 segundos)". Below this, a tip says: "Dica: Tente remover as aspas de sua pesquisa para obter mais resultados." The first result is a PDF document titled "Ferramentas de Busca na Web" from the website pauling.fe.usp.br. The second result is a PDF document titled "MECANISMOS DE BUSCA NA WEB: UMA ANÁLISE DA METODOLOGIA DE ESTUDOS ...".

On the right side of the page, there is a section for "Links Patrocinados" with a link for "Documentos Corporativos" from www.e-docs.com.br and a link to "Veja a sua mensagem aqui...".

intitle:(gravação|cdrom) site:www.dicas-l.unicamp.br



Google™ [Web](#) [Imagens](#) [Grupos](#) [Diretório](#)

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados **1 - 10** dos aproximadamente **13** em **www.dicas-l.unicamp.br** sobre **intitle:(gravação|cdrom) site:www.dicas-l.unicamp.br**

Dicas-L: [Gravação](#) de cds - Kernel 2.6
Enviar esta página por email. **Gravação** de cds - Kernel 2.6. Colaboração ... v
- verbose, mais informações sobre o processo de **gravação**. -eject ...
www.dicas-l.unicamp.br/dicas-l/20040326.php - 6k - [Em cache](#) - [Páginas Semelhantes](#)

Dicas-L: [Como criar o seu proprio CDROM](#) do Linux (2)
... Como criar o seu proprio **CDROM** do Linux (2). ... Uma outra alternativa é você mesmo
gerar a sua imagem, a partir de uma distribuição do Linux em **CDROM**. ...
www.dicas-l.unicamp.br/dicas-l/19990914.shtml - 6k - [Em cache](#) - [Páginas Semelhantes](#)

Dicas-L: [Como criar o seu próprio CDROM](#) do Linux (1)

intitle:(gravação cdrom) site:www.dicas-l.unicamp.br

Google Web [Imagens](#) [Grupos](#) [Diretório](#)

[Pesquisa avançada](#)
[Preferências](#)

Pesquisar: a web páginas em português páginas do Brasil

Web Resultados **1 - 3** dos **3** em **www.dicas-l.unicamp.br** sobre **intitle:(gravação cdrom)**. (9.90 s)

Dicas-L: [Gravação](#) de [CDROM](#) em GNU/Linux
Enviar esta página por email. **Gravação** de **CDROM** em GNU/Linux. Colaboração: Fábio Mengue - UNICAMP. Para gravar um **CDROM** em sistemas ...
[www.dicas-l.unicamp.br/dicas-l/20030117.shtml](#) - 4k - 6 mai. 2004 - [Em cache](#) - [Páginas Semelhantes](#)

Dicas-L: Testar a [gravação](#) de CDs
... Internet e gera os CDs num gravador, após a **gravação** do CD ... a partir do arquivo
4.6.2-disc1.iso: \$ cat /dev/**cdrom** | md5sum bade804f9564165e57647511362a5c96 ...
[www.dicas-l.unicamp.br/dicas-l/20020829.shtml](#) - 4k - [Em cache](#) - [Páginas Semelhantes](#)

[\[Dicas-L\] UNICAMP - CENTRO DE COMPUTAÇÃO](#)
... **Gravação** de **CDROM** em GNU/Linux; Organizando o /usr/local/ com o
GNU/Linux. Ferramentas para o sistema de arquivos. Ferramentas de rede.

Google Images



[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) ^{New!} [more »](#)

~big

Search

[Advanced Image Search](#)
[Preferences](#)

Moderate SafeSearch is on

Images

Results 21 - 40 of about 2,310,000 for ~big. (0.18 seconds)

Show: [All sizes](#) - [Large](#) - [Medium](#) - [Small](#)



fireballmiller-**big**.jpg

1024 x 669 pixels - 48k

[racinginc.com/
images/fireballmiller-big.jpg](http://racinginc.com/images/fireballmiller-big.jpg)



pitch-**big**.jpg

512 x 384 pixels - 26k

[www.sport.ox.ac.uk/
facilities.shtml](http://www.sport.ox.ac.uk/facilities.shtml)



xrad-**big**.gif

910 x 928 pixels - 25k

[www.acm.org/.../videos/
Lamping/xrad-big.gif](http://www.acm.org/.../videos/Lamping/xrad-big.gif)



mw-ntk-**big**.jpg

720 x 504 pixels - 600k

[www.mathworks.com/company/
images/mw-ntk-big.jpg](http://www.mathworks.com/company/images/mw-ntk-big.jpg)

Google Images

Busca Avançada



Advanced Image Search

[Images Help](#) | [About Google](#)

Find results	related to all of the words	<input type="text" value="linus torvalds initial linux announ"/>	<input type="button" value="Google Search"/>
	related to the exact phrase	<input type="text"/>	
	related to any of the words	<input type="text"/>	
	not related to the words	<input type="text"/>	
Size	Return images that are	<input type="text" value="any size"/>	<input type="button" value="any size"/>
Filetypes	Return only image files formatted as	<input type="text"/>	<input type="button" value="small"/>
Coloration	Return only images in	<input type="text"/>	<input type="button" value="medium"/>
Domain	Return images from the site or domain	<input type="text"/>	<input type="button" value="large"/>
SafeSearch	<input type="radio"/> No filtering <input checked="" type="radio"/> Use moderate filtering <input type="radio"/> Use strict filtering		

Google Groups



[Web](#) [Images](#) [Groups](#) [News](#) [Froogle](#) ^{New!} [more »](#)

linus torvalds initial linux announcement

Search

[Advanced Groups Search](#)
[Preferences](#)

Groups

Results 1 - 100 of about 191 English and Portuguese messages for [linus tor](#)

Related groups: [fido7.ru.linux.*](#) (2 groups)
[comp.os.linux.announce](#)
[comp.os.linux.*](#) (20 groups)
[comp.os.linux.misc](#)

[\[1/2\] Amigactive Magazine Dissects Amiga's Linux Announcement](#)

... eight months since the **initial** partnership **announcement**. ... open source traditions and philosophies of **Linux**. Amigactive Magazine asked **Linus Torvalds**, author of ...
[fido7.ru.amiga](#) - 14 Jul 1999 by RDC - [View Thread \(1 article\)](#)

[Re: The shame of Linux](#)

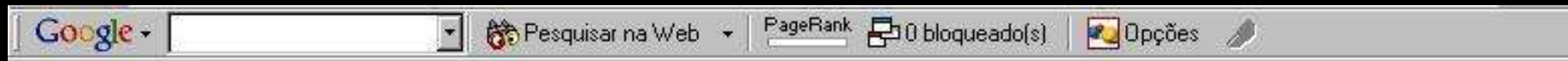
... format ----- N: **Linus Torvalds** E: **Linus.Torvalds@Helsinki.FI** D ... recent disruption of the **Linux**-Activists KERNEL ... Both **initial** submissions and updates continue to ...
[comp.os.linux.misc](#) - 16 Dec 1994 by John Martin - [View Thread \(27 articles\)](#)

[Linux/Alpha Progress Report - Part 2](#)

... He provided the loaner system to **Linus Torvalds**. ... I'd like to thank YOU, the **Linux** community, for ... outpouring of support following my **initial announcement** of the ...
[comp.os.linux.announce](#) - 18 Nov 1994 by Jim Paradis - [View Thread \(1 article\)](#)

Google Toolbar

<http://toolbar.google.com>



Felicitações!

Agora que você instalou o Google Toolbar, você possui a melhor ferramenta de busca do mundo em suas mãos.

Começando

Se algum outro browser do windows do Internet Explorer estiver ainda aberto, feche eles agora.



1. Digite seus termos de busca na caixa de busca.
2. Pressione Enter ou clique no botão search para encontrar resultados. Clique na seta para baixo para selecionar entre diferentes métodos de busca.

Para detalhes completos de como usar e customizar sua Barra de Ferramentas Google, [clique aqui](#).

Google News

- O Google oferece um serviço de notícias, o **Google News (<http://news.google.com>)**, que varre milhares de sites de notícias e monta automaticamente uma página, onde as notícias selecionadas são exibidas.

Search and browse 4,500 news sources updated continuously.

Top Stories

[World](#)[U.S.](#)[Business](#)[Sci/Tech](#)[Sports](#)[Entertainment](#)[Health](#) [News Alerts](#)[Text Version](#)[About](#)

Top Stories

Auto-generated 15 minutes ago

[Iraq Prisons Chief Says Abu Ghraib To Remain Open](#)



BBC News

Voice of America -
1 hour ago

The head of US-run prisons in Iraq says the military will continue operating the Baghdad prison where abuse of detainees by American soldiers recently became public.

[Soldier describes unclear guidance at Abu Ghraib prison](#)

San Jose Mercury News (subscription)

[No Plans to Close Abu Ghraib, American Prison Commander Says](#)

New York Times

Guardian Boston Globe Channel

['Racketeer' Black faces ruin](#)

Guardian - [and 193 related »](#)

[Intel Scraps Pentium 4, Moves to Dual Core Chip](#)

NewsFactor Network - [and 239 related »](#)

[Oberholser Turns It Up as Woods Feels the Heat](#)

Washington Post - [and 716 related »](#)

[Friends We Must Leave Behind](#)

Sunday Herald - [and 876 related »](#)

[Martin's doctor runs private clinic](#)

Canada.com - [and 33 related »](#)

Google News: Alertas



Associado ao serviço **Google News**, existe um serviço de notificação, que permite que sempre que aparecer uma manchete associada a um determinado assunto sobre o qual você queira se manter informado, um email seja enviado para o endereço que você especificar.

<http://www.google.com/newsalerts>

A Web Oculta



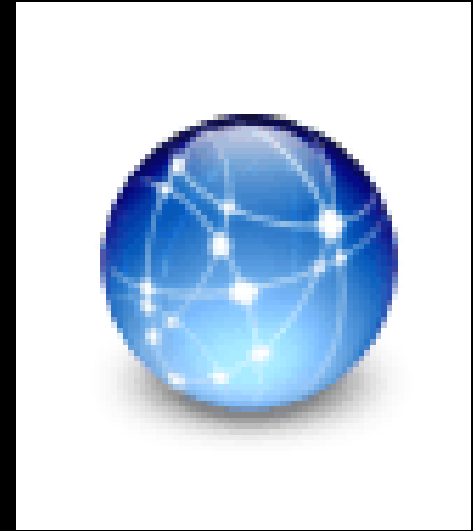
- Também conhecida como a web profunda
- A maior parte da informação na web é gerada dinamicamente, através de consulta a bancos de dados
- Mecanismos de busca normais não alcançam esta informação

A Web Oculta



- Os mecanismos de busca constroem seus índices navegando de página em página (spidering) através da internet
- Para ser encontrada, uma página precisa ser apontada por outras
- Mecanismos de busca não conseguem recuperar informações armazenadas em bancos de dados
-

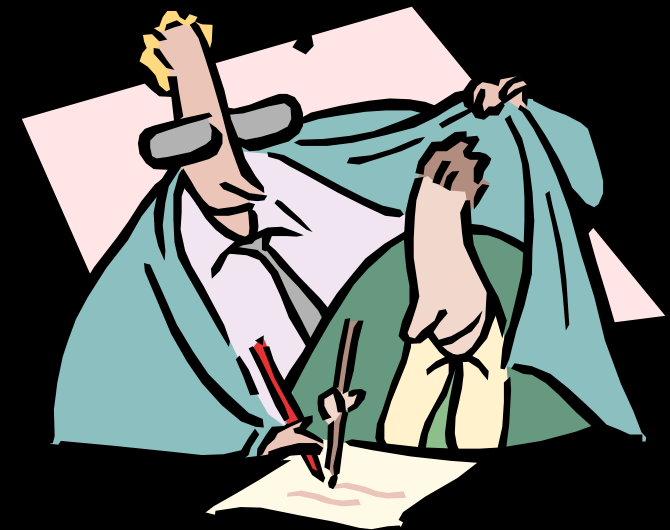
O tamanho da Web Oculta



- Segundo um estudo de março de 2000:
 - A informação contida na web oculta é 400 a 550 vezes maior que a da web pública
 - Categoria de informação que mais cresce na Internet
 - Maior qualidade

A Web Oculta

- Os maiores indexadores (Google , Northern Light etc.) indexam apenas 16% da web.
- Buscas na web com os mecanismos mais conhecidos pesquisam apenas em 0.03% do conteúdo existente



Vortais

Mecanismos de Busca Especializados

- National Cancer Institute
 - <http://www.cancer.gov>
- Medical World Search
 - <http://www.mwsearch.com/>
- Computer Science Research Papers
 - <http://citeseer.ist.psu.edu/cis>
- MedHunt
 - <http://www.hon.ch/MedHunt/>
- Specialized Search Engines
 - <http://www.searchability.com/>
- Biblioteca Digital da Unicamp
 - <http://libdigi.unicamp.br>



Índice

[Página principal](#)

[Documentos](#)

[Usuários](#)

Procurar por:

[Procura avançada](#)

[Dúvidas e sugestões](#)

Introdução

Bem-vindo à **Biblioteca Digital da Unicamp**, que utiliza o sistema **Nou-Rau**, implementando um serviço online para armazenamento e obtenção de documentos, provendo acesso controlado e mecanismos eficientes para busca.

Totais

Documentos: 6273

Teses: 1909

Novidades

Tópicos principais

Tópicos

[Congressos e Seminários](#) Trabalhos apresentados em eventos

[Dissertações e Teses](#) Produção Científica da UNICAMP

[Hemeroteca - CMU - Campinas](#) Acervo Digital

[Periódicos Eletrônicos UNICAMP](#) Periódicos publicados pela UNICAMP em Formato Eletrônico

[Teses mais acessadas](#)

[Outros documentos mais acessados](#)



Visitante
Ler/Perguntar
Ranking de Colaboradores
Estatísticas

Bem-vindo ao Rau-Tu

O **Rau-Tu** é um sistema que cobre diversas áreas de conhecimento, denominadas tópicos. Para formular perguntas, assim como acessar o repositório de perguntas e suas respectivas respostas, não é necessário se cadastrar.

Para participar do **Rau-Tu** respondendo perguntas formuladas pelos visitantes você deverá [cadastrar-se como colaborador](#). Mais informações sobre o **Rau-Tu** leia em [projeto rau-tu](#) e [ajuda](#).

Colaboradores em destaque
2866 humberto
1611 infernet
1402 eduardok
1312 jocajuni
768 jeveaux

[Veja o ranking completo](#)

Tópicos Disponíveis

Tópicos

Ambientes Gráficos	KDE, GNOME, gerenciadores de janelas e servidor X
Audio	Som no Linux, MP3, sintetizadores, configuração
Distribuições	Distribuições do Linux: RedHat, Debian, Conectiva, SuSE, Slack...
Hardware	Compatibilidade e configuração de hardware no Linux
Jogos	Jogos em geral, instalação, configuração
Kernel	Kernel hacking, compilação, tuning, etc...
Networking	Administração, DNS, Sendmail, Samba, Redes em geral
Outras Arquiteturas	Linux rodando em Sparc, Alpha, PowerPC, etc...
Programação	Desenvolvimento no Linux com C, C++, PHP, Java, Perl, etc...
Segurança	Questões de segurança no Linux

Novidades

Atualização Versão Rau-Tu

enviada por [admin](#) em 12/08/2003 16:11

O Rau-Tu de Linux está usando agora a versão 1.2.1, que contém diversas novas funcionalidades, principalmente do ponto de vista do administrador, e que apresenta também um



Índice

[Página principal](#)

[Documentos](#)

[Novidades](#)

[Usuários](#)

Procurar por:

Procurar

[Procura avançada](#)

Introdução

Bem-vindo ao sistema **Nou-Rau**, que implementa um serviço online para armazenamento e obtenção de documentos, provendo acesso controlado e mecanismos eficientes para busca.

Totais

Documentos: 67

Tamanho total: 96 Mb

Novidades

Read the notice [archive](#)

Tópicos principais

Tópicos

[Advocacia](#) Documentos para defesa e divulgação do software livre e aberto

[Banco de Dados](#) Banco de dados Open Source

[Ferramentas Gráficas](#) Ferramentas de Edição Gráficas

[Ferramentas de Escritório](#) Processadores de Texto, Planilhas Eletrônicas, etc

[Internet](#) Ferramentas de Navegação Livre

[Palestras e Apresentações](#) Material de palestras sobre software livre

[Programação](#) Ferramentas de Programação

[Segurança](#) Documentos relacionados a segurança de sistemas e aplicações

[Sistemas Operacionais](#) Sistemas Operacionais Livres

Documentos mais acessados

Downloads

31469 [Linux Total](#)

6093 [Guia Foca Linux Avançado](#)

5722 [Avaliação Técnica do MySQL](#)

5627 [Descrição dos Aplicativos do Conectiva 8](#)

4149 [tutor MySQL](#)

Metabuscadores

- Não indexam a web
- Usam os recursos de outros indexadores
 - Enviam as perguntas, formatadas segundo a sintaxe de cada um deles
 - Recebem e combinam os resultados
 - Apresentam uma lista ao usuário
- Valor maior reside na melhor combinação dos resultados

Metabuscadores

- Profusion
 - <http://www.profusion.com>
- DogPile
 - <http://www.dogpile.com>
- Vivissimo
 - <http://vivisimo.com/>
- HotBot
 - <http://www.hotbot.com>
- Mamma
 - <http://www.mamma.com>



Mapas registados

Novo mapa

Temas :

- recursos humanos
- treinamento e desenvolvimen
- empresa
- necessidades
- habilidades
- cria
- educação
- prática
- prof
- suas
- seres
- recursos
- conhecimento
- centro
- palestras
- neste
- treinamento
- particulares
- comportamento
- máximo
- modelo
- nova
- pessoas

Buscas frequentes

Últimas buscas

Búsquedas frequentes



Dicas de Busca

- Prefira frases, com os termos de busca próximos uns aos outros:
- “medicina pediátrica” e não “medicina” e “pediátrica”
- Use tantos sinônimos quanto puder – os mecanismos de busca utilizam métodos estatísticos para recuperação da informação e os resultados são melhores quando são fornecidas mais palavras

Dicas de Busca (...)

- Evite usar palavras muito comuns, como “computador”, “internet”, etc.
- Entre os argumentos de busca em letras minúsculas. Use as maiúsculas para forçar uma busca exata (não se aplica ao Google)
- Use a opção ‘More like this’ option, se suportada pelo mecanismo de busca (Google, Excite) Google)

***Comunicação
global***



***Poder para
pessoas comuns
A força das idéias
Solidariedade
Invencível***



Referências

- The Deep Web: Surfacing Hidden Value
<http://www.brightplanet.com/technology/deepweb.asp>
- Search Engine Watch
<http://www.searchenginewatch.com>
- A Leitura nos Oceanos da Internet
Editora Cortez
- About Web Search - Guide to Search Engine Optimization & Online Search -
<http://websearch.about.com>



Referências (...)

- Internet Society (ISOC) All About The Internet: History of the Internet - <http://www.isoc.org/internet/history>
- Dave's Site - History of the Internet
<http://www.davesite.com/webstation/net-history.shtml>
- Web Directory – definição, informação, sites, artigos http://www.marketingterms.com/dictionary/web_directory
- Busca na Web
<http://www.ead.unicamp.br/minicurso/bw/index.html>



Referências (...)

- Webopedia: The Difference Between the Internet and the World Wide Web -
http://www.webopedia.com/DidYouKnow/Internet/2002/Web_vs_Internet.asp
- Sink or Swim: Internet Search Tools & Techniques -
<http://www.ouc.bc.ca/libr/connect96/search.htm>
- DARPA – Transformation and Transition: DARPA’s Role in Fostering an Emerging Revolution in Military Affairs Volume 1 – Overall Assessment -
<http://www.darpa.mil/body/newsitems/pdf/idarmastudyvol1.pdf>



Referências (...)

- **O Buscador dos Buscadores**
<http://www.finderseeker.com>
- **Virtual Free Sites**
<http://www.virtualfreesites.com>



Contato

Rubens Queiroz de Almeida

queiroz@unicamp.br

<http://www.Dicas-L.unicamp.br/cursos>