

Apresentação da Disciplina

Prof. Dr. Leandro Balby Marinho

<http://www.dsc.ufcg.edu.br/~lbmarinho>



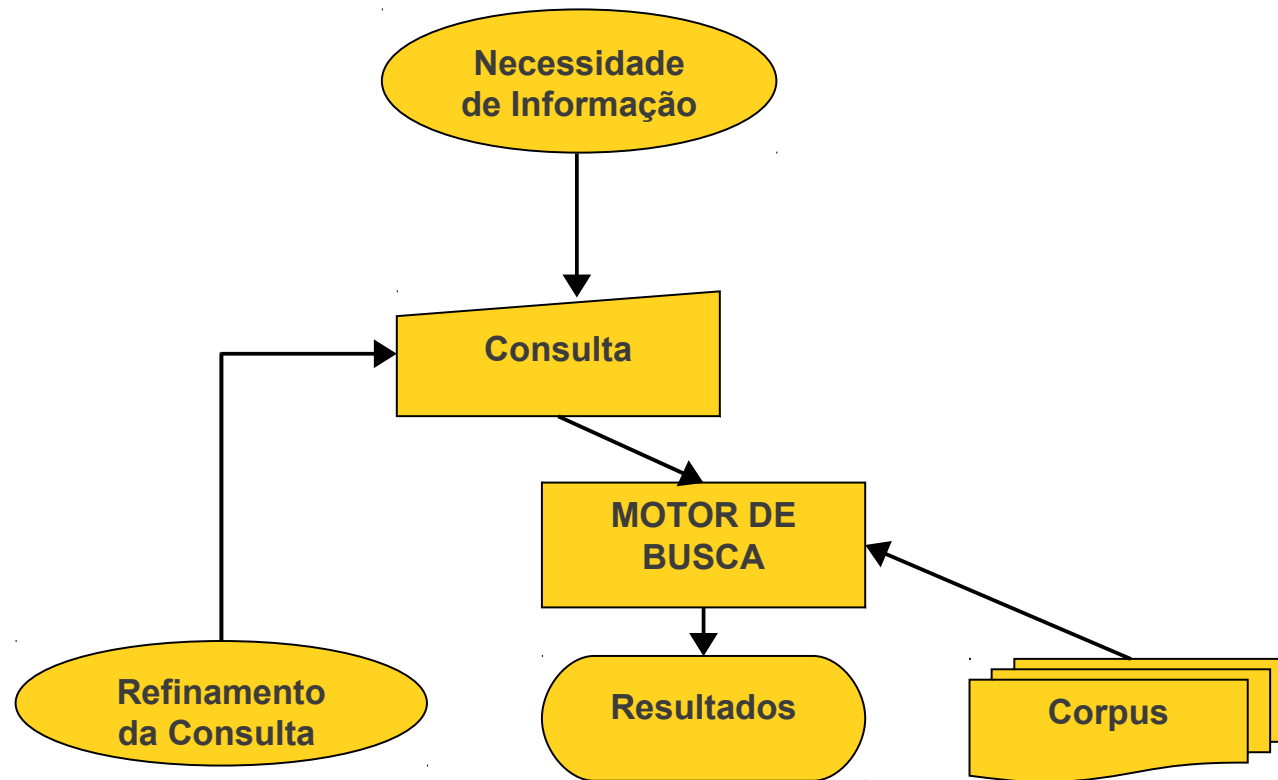
UFCG CEEI Departamento de
Sistemas e
Computação

Sistemas de Recuperação da Informação

Recuperação da Informação (RI)

- O que é RI?
- Por que RI?

Recuperação da Informação (RI)



Antes...



Hoje...



Exemplo

- Quais peças de Shakespeare contém **Brutus AND César** mas **NOT Calpurnia**?
- Poderia-se procurar em todas as peças de Shakespeare contendo **Brutus AND César**, e excluindo as peças contendo **Calpurnia**
- Por que não é uma boa solução?
 - Inviável para grandes coleções de documentos
 - Pode-se querer outras operações (e.g., achar a palavra **romanos** perto de **compatriota**) não é viável
 - Como lidar com ambiguidade?
 - Pode-se querer a ordenação dos resultados

Sistemas de RI podem recuperar...

Web [Imagens](#) [Vídeos](#) [Mapas](#) [Notícias](#) [Orkut](#) [Gmail](#) [mais](#) ▼



UFMG



Pesquisar

Aproximadamente 243.000 resultados (0,32 segundos)

[Pesquisa avançada](#)

- Tudo
- Imagens
- Vídeos
- Notícias
- Mais

Campina Grande - PB
[Alterar local](#)

A Web

[Páginas em português](#)
[Páginas de Brasil](#)
[Páginas estrangeiras traduzidas](#)

Em qualquer data

[Mais recentes](#)
[Últimas 24 horas](#)
[Última semana](#)
[No último mês](#)
[Último ano](#)
[Intervalo personalizado...](#)

[Mais ferramentas](#)

Portal da Universidade Federal de Campina Grande - PB

A **Universidade Federal de Campina Grande** divulgou na tarde desta sexta, dia 11, a terceira chamada de vagas remanescentes do Vestibular 2011 da instituição. ...

www.ufcg.edu.br/ - [Em cache](#) - [Similares](#)

[Comprov](#)

[PB](#)

[UFMG/Universidade Federal de ...](#)

[Mais notícias](#)

[Cursos](#)

[Calendário escolar](#)

[Pró-Reitoria de Pós-Graduação](#)

[Cursos de especialização ano 2006 ...](#)

[Mais resultados de ufcg.edu.br »](#)



Ufcg ☆ 🔍

[Página do local](#)

Rua Aprígio Veloso, 882 - Bodocongó
Campina Grande - PB, 58109-900
(0xx)83 3310-1467

★★★★★ 5 resenhas - [Escrever resenha](#)

Notícias

Joomla! - O sistema dinâmico de portais e gerenciador de conteúdo.
Vestibular 2011 - Vestibular 2010 - Residência Médica 2011 - Concursos
www.comprov.ufcg.edu.br/ - [Em cache](#) - [Similares](#)

COMPROV - Notícias

Resolução Nº 12/2010 (Número de vagas do Vestibular 2011 da **UFMG** ...
www.comprov.ufcg.edu.br/index.php?option=com... - [Em cache](#)

... documentos de texto

Sistemas de RI podem recuperar...

[Web](#) [Imagens](#) [Vídeos](#) [Mapas](#) [Notícias](#) [Orkut](#) [Gmail](#) [mais ▼](#)

ibalby@gmail.com | [Configurações ▼](#) | [Se](#)

Google

UFCG

Pesquisar

SafeSearch moderado ▼

Aproximadamente 31.500 resultados (0,40 segundos)

[Pesquisa avançada](#)

[Tudo](#)

[Imagens](#)

[Vídeos](#)

[Notícias](#)

[Mais](#)

Qualquer tamanho

Grande

Médio

Ícone

Maior que...

Exatamente...

Qualquer tipo

Rosto

Foto

Clip art

Desenho linear

Qualquer cor

Colorida

Preto-e-branco



Visual. padrão

Mostrar tam



... imagens.

Sistemas de RI podem recuperar...

[Web](#) [Imagens](#) [Vídeos](#) [Mapas](#) [Notícias](#) [Orkut](#) [Gmail](#) [mais ▼](#)



UFMG

Pesquisar

SafeSearch moderado ▼

Aproximadamente 1.640 resultados (0,10 segundos)

[Pesquisa avançada](#)

Tudo

Imagens

Vídeos

Notícias

Mais

Todas as durações

Curto (0-4 min)

Médio (4-20 min)

Longo (20 min ou mais)

Em qualquer data

Última hora

Últimas 24 horas

Última semana

No último mês

Último ano

Intervalo personalizado...

Classificados por relevância

Classificados por data

Qualquer qualidade

Alta qualidade

Todos os vídeos



Aula dos Feras - UFCG - Computação 2009.1 ☆

4 min - 9 mar. 2009 - Vídeo enviado por MatheusSp
Aula dos feras do curso de computação. Cívico :)
[youtube.com](#) - [Vídeos relacionados](#)



DSC em pauta: Computação / UFCG - 05 Estrelas ☆

2 min - 26 set. 2007 - Vídeo enviado por sergiodbe
Reportagem destaca conceito máximo em ranking nacional. Dois cursos da **UFMG** (Ciência da Computação e Matemática ...
[youtube.com](#) - [Vídeos relacionados](#)



Pyron Man - UFCG Proglab1 2010.1 ☆

6 min - 1 jul. 2010 - Vídeo enviado por wcp1991
Jogo de plataforma produzido por Dalton Dinarte e Wallisson de Carvalho.
[youtube.com](#) - [Vídeos relacionados](#)



Calourada UFCG - VT incompleto ☆

30 seg - 25 mar. 2007
Calourada é banda voou mesmo, daí... o VT não saiu...
[videolog.tv](#) - [Vídeos relacionados](#)



YouTube - computacaoufcg Channel: Ciência ... ☆

Vídeos relacionados ao Curso de Ciência da Computação da Universidade Federal de Campina Grande (UFCG)

... vídeos.

Sistemas de RI podem recuperar...

[jobs](#) [salaries](#) [trends](#) [forums](#)



what:

Information Retrieval

job title, keywords or company

where:

Find Jobs

[Advanced Job Search](#)

Information Retrieval jobs

Jobs 1 to 10 of 7,741

▼ Salary Estimate

\$20,000+ (6891)
\$40,000+ (4685)
\$60,000+ (3112)
\$80,000+ (1823)
\$100,000+ (854)

► Title

► Company

► Location

► Job Type

► Employer/Recruiter

► [Get new jobs for this search by email](#)

Tip: Enter your zip code in the "where" box to show results in your area.

Sort by: **relevance** - [date](#)

[Senior Software Engineer- Information Retrieval an...](#)

Amazon - Seattle, WA

a variety of **Information Retrieval** and Data Mining... of **Information** present in Amazon and solving targeted problems. Incorrect or missing **Information** for...

From Amazon - 4 days ago - [save job](#) - [block](#) - [email](#) - [more...](#)

[Needle Machine Learning/Information Retrieval Inte...](#)

ITA Software - United States

Machine Learning/**Information Retrieval** Internship... Skills: Experienced in Machine Learning and **Information Retrieval** Proven development experience in Java...

From ITA Software - 13 days ago - [save job](#) - [block](#) - [email](#) - [more...](#)

[Caseworker \(Medical Records Retrieval\)](#)

Examination Management Services, Inc. - Omaha, NE - [+1 location](#)

insurance applicants by calling medical facilities Verifying the applicant **Information** Negotiating the medical fees and the expedition of the records Monday...

From Examination Management Services, Inc. - 5 days ago - [save job](#) - [block](#) - [email](#) - [more...](#)

[Video Tech Retrieval Specialist](#)

Bank Of America - San Francisco, CA

processed in the Photo **Retrieval** Datga Base. Must be... knowledge of **Information** Privacy Protection and how to encrypt and safe guard **Information**. Must be willing...

From Bank of America - 11 days ago - [save job](#) - [block](#) - [email](#) - [more...](#)

Email this search to a friend:

From my email address

To email address

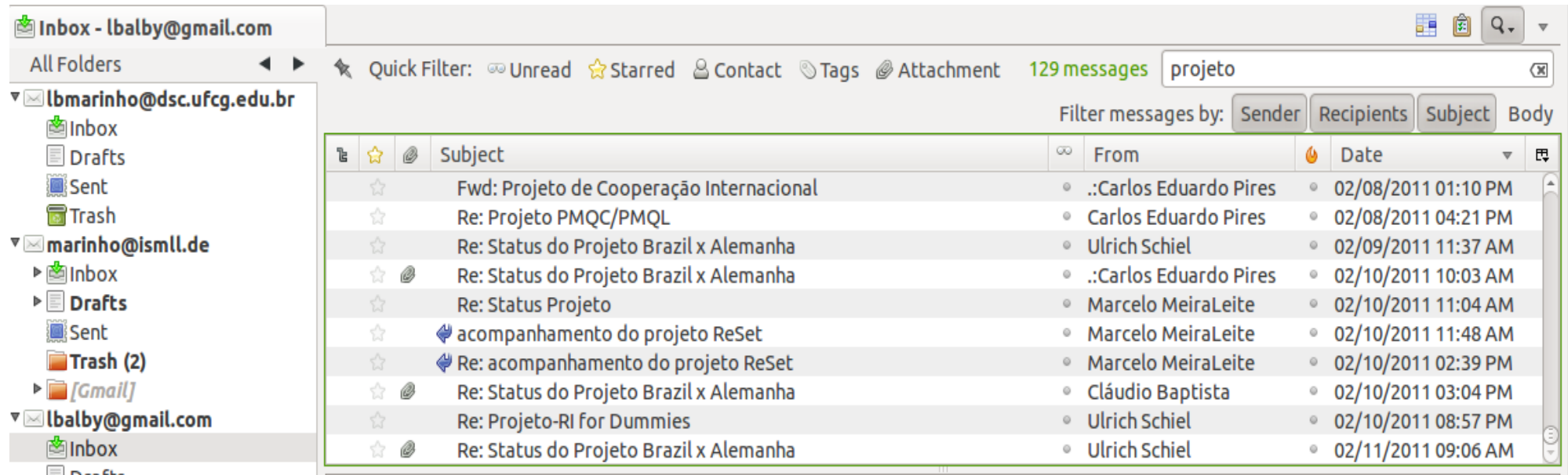
► [Add a message \(optional\)](#)

Send

[RSS Job Feed](#)

... empregos entre muitos outros.

Sistemas de RI também estão no ...



... email.

Sistemas de RI também estão nos ...

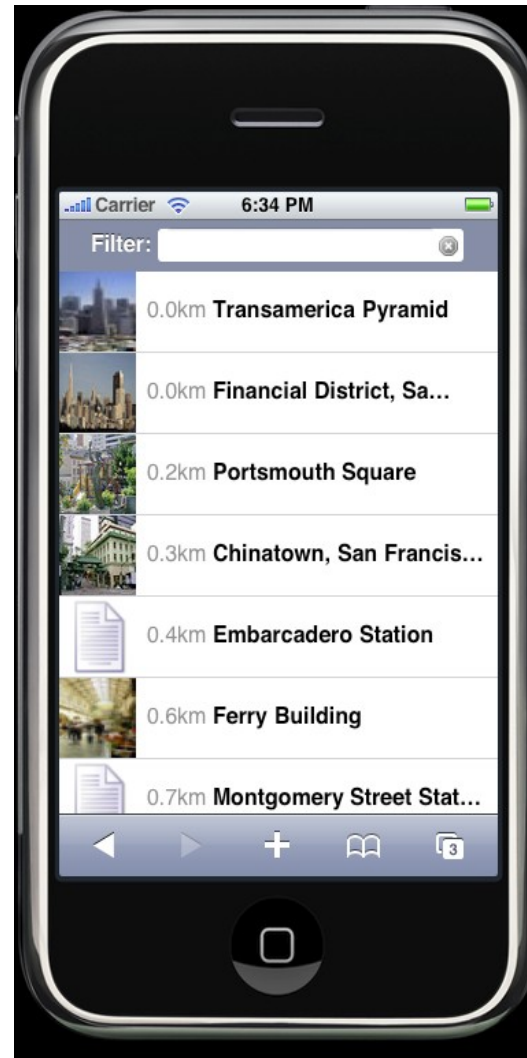


Note: partial results only (34.7% complete). One-time index update in progress.

-  [apresentacao_disc.odp](#)
 ... Dr. Leandro Balby Marinho **Apresentação** da Disciplina Prof. Dr. Leandro Balby Marinho
<http://www.dsc.ufcg.edu.br/~lbmarinho> Sistemas de Recuperação da Informação Corpus Necessidade
 de Informação Consulta Resultados MOTOR DE BUSCA Refinamento da Consulta Antes... Hoje...
 Recuperação da ...
/home/leandro/Documents/DSC/cursos/.../apresentacao_disc.odp - [Open folder](#) - [3 cached](#) - 02/14/2011
-  [OpenOffice.org Presentation](#)
 Presentation - Create and edit presentations for slideshows, meeting and Web pages.
</usr/share/applications/openoffice.org-impress.desktop> - [Open folder](#) - 02/14/2011
-  [aula2-dicionario.odp](#)
 ... Com Unicode, a **apresentação** é complexa na superfície, mas o formato de armazenamento is
 simples. s Complicações de Idioma Cria-se uma lista contendo as palavras mais frequentes, de forma
 a excluí-las do dicionário. Intuição: Possuem fraco conteúdo semântico: a , o, e, para, ser Há ...
</home/leandro/Documents/DSC/cursos/R.../aula2-dicionario.odp> - [Open folder](#) - [1 cached](#) - 02/04/2011
-  [Prof. Dr. Leandro Balby Marinho](#)
 ... Teaching Projects Other **Apresentação** Slides A Recuperação da Informação (RI) trata da busca por
 documentos, informações dentro desses documentos e metadados sobre esses documentos, em
 grandes coleções de dados desestruturados (e.g., documentos da Web). Muito embora a área não seja
 ...
/home/leandro/Documents/homepage/ri_apres_2011.1.html - [Open folder](#) - [1 cached](#) - 02/03/2011
-  [Prof. Dr. Leandro Balby Marinho](#)
 ... Teaching Projects Other **Apresentação** Slides Matemática discreta trata do estudo de estruturas
 matemáticas discretas, e.g. álgebra Booleana, grafos, relações e outras, enquanto outras

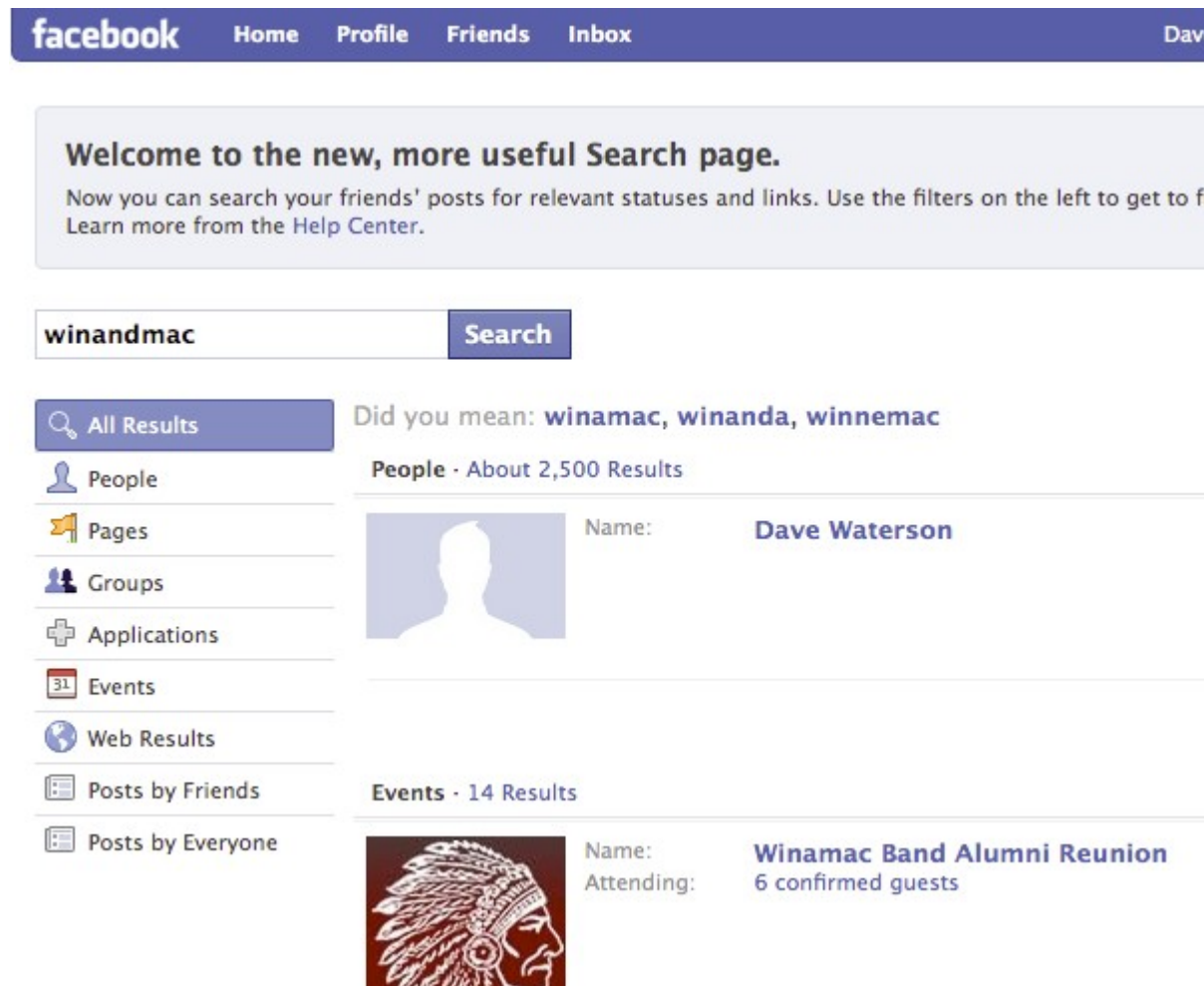
... desktops.

Sistemas de RI também estão nos ...



... dispositivos móveis.

Sistemas de RI também estão nos ...



... sites de relacionamento.

Sistemas de RI também estão nas ...



CAPES
Universidade Federal da Paraíba

[SIGN IN](#) [SIGN UP](#)

SEARCH

Full text of every article ever published by ACM and bibliographic citations from major publishers in computing.

- [Using the ACM Digital Library](#)
- [For Consortia Administrators](#)

Announcements

- **ACM Transactions the Web (TWEB) Seeking Editor-in-Chief**
Nominations are due **April 1, 2011**. For more information please visit the [TWEB EIC search page](#).
- **New publication: ACM Transactions on Management Information Systems (TMIS)**
Read more about *TMIS* on the [home page](#) and in the [Digital Library](#).

Recently loaded issues and proceedings:

(available in the DL within the past 2 weeks)

ACM Journal on Emerging Technologies in Computing Systems (JETC)

[Volume 7 Issue 1](#)

Journal of Data and Information Quality (JDIQ)

[Volume 2 Issue 2](#)

ACM Transactions on Programming Languages and Systems (TOPLAS)

[Volume 33 Issue 2](#)

Queue

[Volume 9 Issue 2](#)

Advanced Search

Browse the ACM Publications:

- [Journals](#)
- [Transactions](#)
- [Magazines](#)
- [Newsletters](#)
- [Special Interest Groups \(SIGs\)](#)
- [Proceedings](#)
- [ACM Oral History interviews](#)

Browse the Conferences:

- [Recent and Upcoming Conferences](#)
- [Conference Listing](#)

Browse the [Publications by Affiliated Organizations](#)

Browse all literature by type

Browse all literature by [Publisher](#)

Browse by the [ACM Computing Classification System](#)



[The Encyclopedia of Computer Science](#)
"...is the definitive reference in computer science and technology..."

... bibliotecas digitais.

Sistemas de RI também estão nas ...



... tvs entre outros.

Dados Estruturados vs. Desestruturados

- Dados estruturados normalmente se referem a informações em “tabelas” de bd relacionais.

Empregado	Gerente	Salario
Smith	Jones	50000
Chang	Smith	60000
Ivy	Smith	50000

Normalmente permite escala numérica e casamento exato de consultas textuais, e.g., *Salario < 60000 AND Gerente = Smith*.

Dados Desestruturados

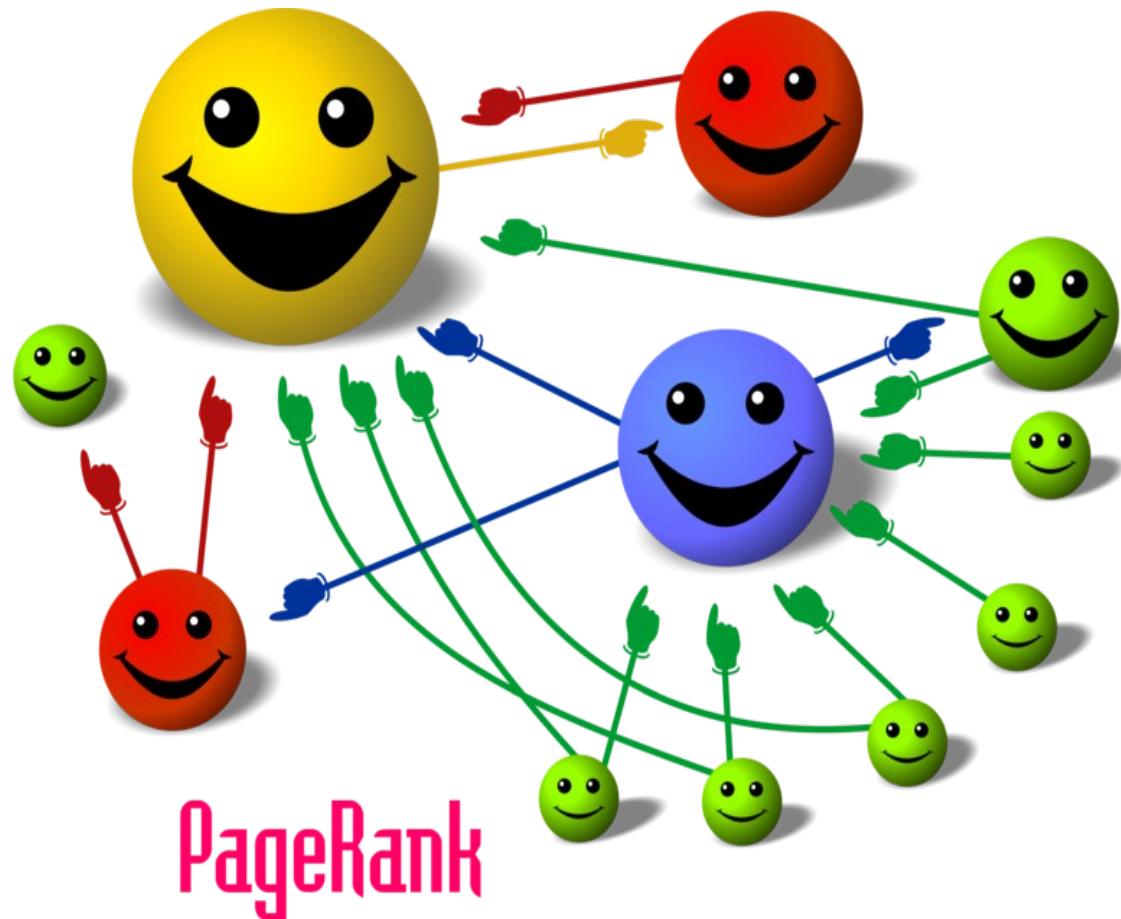
- Tipicamente referem-se a texto livre.
- Permitem
 - Consultas de palavras-chave incluindo operadores (e.g., AND, OR, NOT)
 - Consultas mais sofisticadas de “conceitos” e.g.,
 - ache todas as páginas da web que tratam de abuso de drogas

Dados Semi-Estruturados

- Na verdade quase nenhum dado é inteiramente “desestruturado”.
- E.g., esse slide possui zonas distintas tais como *Título* e *Marcadores*.
- Facilita busca “semi-estruturada” tais como
 - *Título contém dados AND Marcadores contém busca*

... além de estruturas linguísticas latentes.

Análise de Hiperlinks



Agrupamento, Classificação e Ranking

- **Agrupamento:** Documentos são agrupados de acordo com seu conteúdo.
- **Classificação:** Determinado doc D é classificado em um tópico dado um conjunto prédefinido de tópicos.
- **Ranking:** Aprender a melhor forma de ordenar um conjunto de documentos, e.g., um conjunto de resultados de pesquisa.

Modelos mais Sofisticados de RI

- Recuperação de Informação entre Idiomas
- Recuperação Multi-mídia
- Responder a perguntas
- Sumarização
- Mineração em Texto
- Sistemas de Recomendação
- ...

Detalhes do Curso

- URL do curso:
 - [www.dsc.ufcg.edu.br/~lbmarinho/ri_apres_2011.1.html]
- Horário: Seg. (16h) e Quinta (14h), sala: RE08
- Avaliação:
 - Prova Escrita
 - Seminário
 - Mini Projetos
- Livro Texto:
 - *Introduction to Information Retrieval*
 - Disponível online em (<http://informationretrieval.org/>)