

A Geração de Riqueza a Partir da Universidade: o Caso Akwan

Nivio Ziviani

Departamento de Ciência da Computação da UFMG

Curso de Recuperação de Informação – Máquinas de Busca na Web

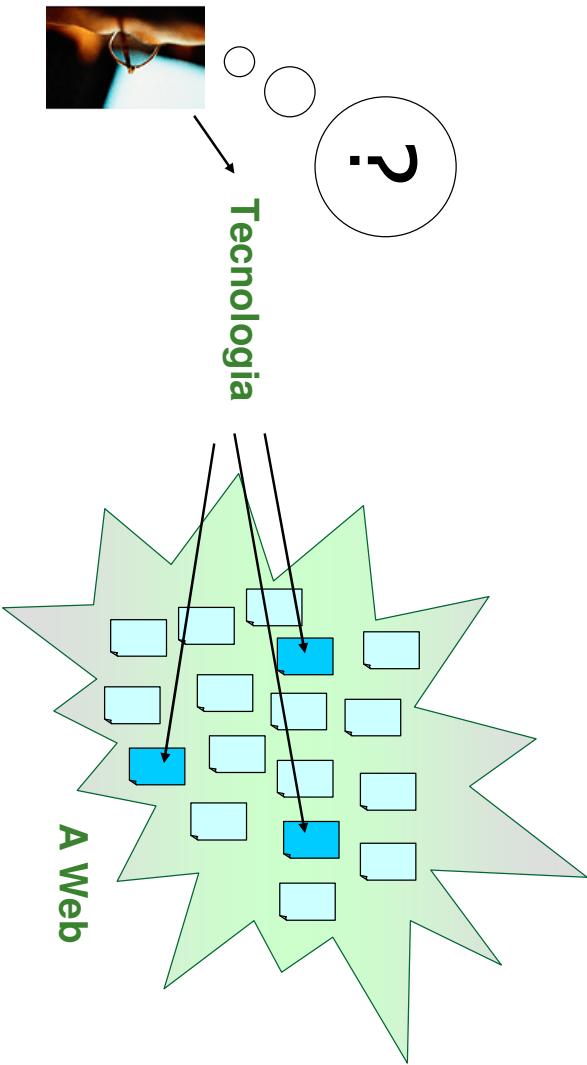
Programa de Pós-Graduação em Ciência da Computação da UFMG

28 de Fevereiro de 2007

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

1

O Problema



Busca na Web: Necessidades dos Usuários

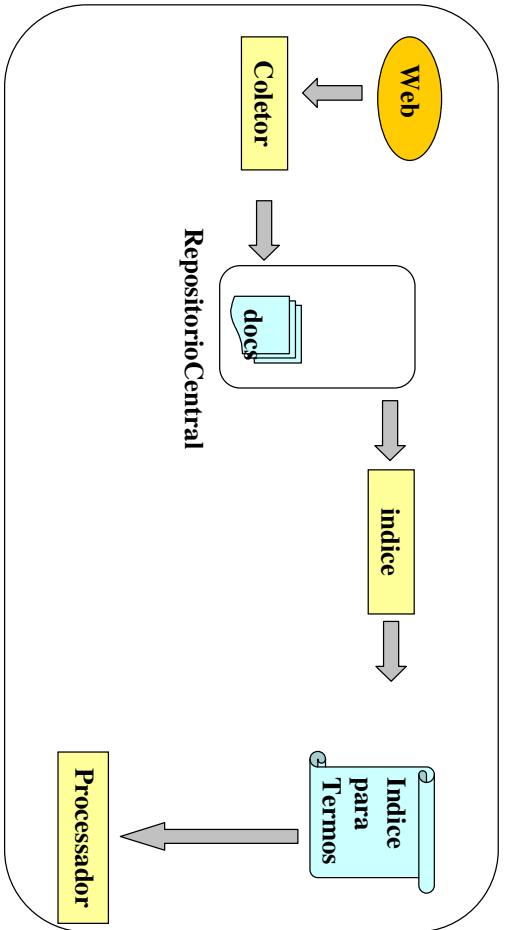
- Coleção: Web pública: estática + dinâmica
- Objetivo: recuperar informação qualificada que seja relevante para as necessidades dos usuários
- Necessidade
 - Informational – deseja **aprender** sobre algo (~40%)
 - Navigational – deseja **ir** para aquela página (~25%)
 - Transactional – deseja **fazer** algo (web-mediated) (~35%)
 - Acessar um serviço (“Tempo em Salvador”)
 - Downloads (“Imagens sobre superfície de Marte”)
 - Compra (“Pentax Optio”)

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

3

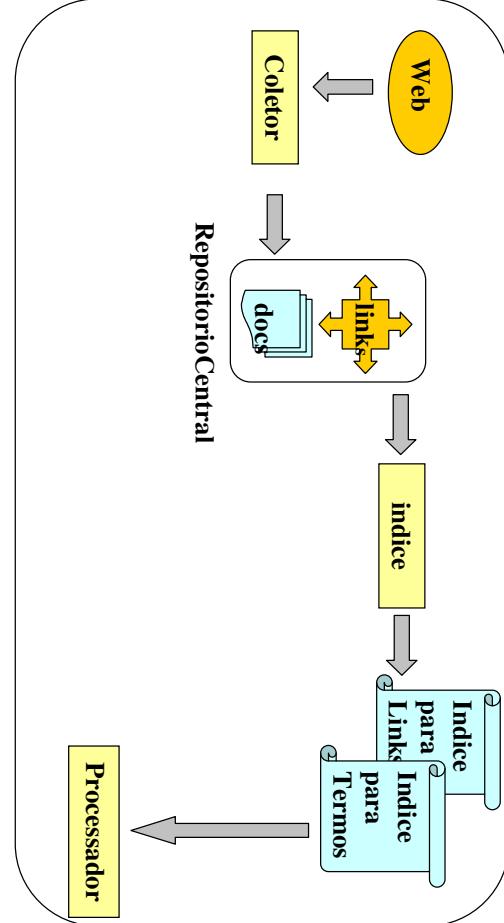
Evolução da Tecnologia de Busca

Primeira Geração



Evolução da Tecnologia de Busca

Segunda Geração

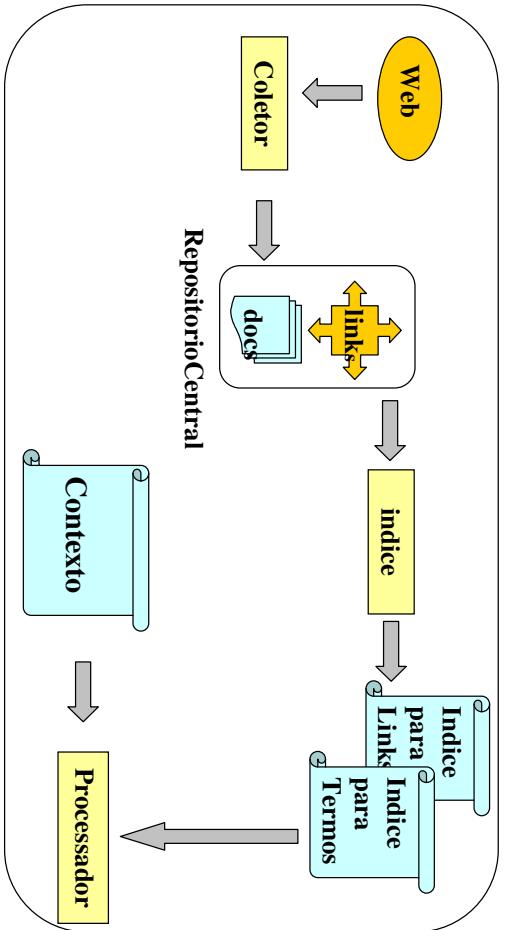


LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

5

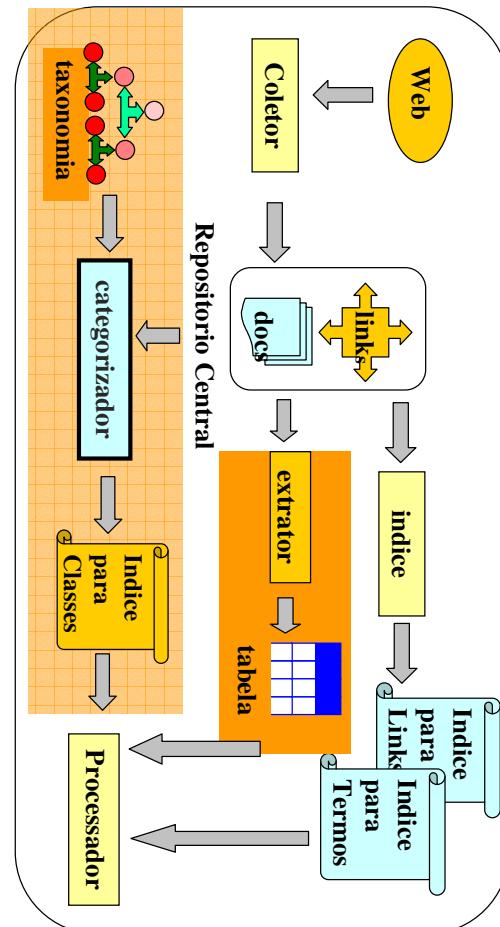
Evolução da Tecnologia de Busca

Terceira Geração



Evolução da Tecnologia de Busca

Busca Corporativa



LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

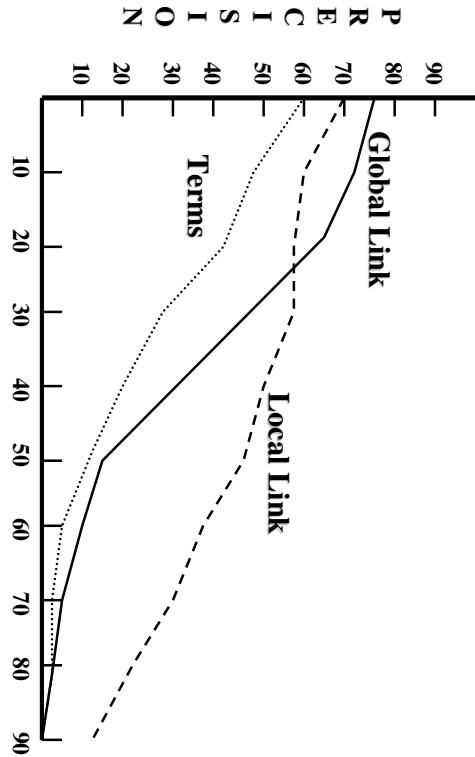
7

Exemplos de Tópicos de Pesquisa

- **Modelos de RI:** determinam a precisão da resposta a consultas e formam a base tecnológica do componente principal de qualquer sistema de RI, o *processador de consultas*
- **Eficiência em RI:** Sistemas de RI, além de eficazes, devem ser extremamente eficientes, uma vez que o *tempo de resposta* é um problema central
- **Categorização em RI:** identificação automática de características de documentos e uso dessas características para determinar se os documentos pertencem ou não a uma categoria específica

Information Retrieval in the Web

Second generation search engine: link analysis



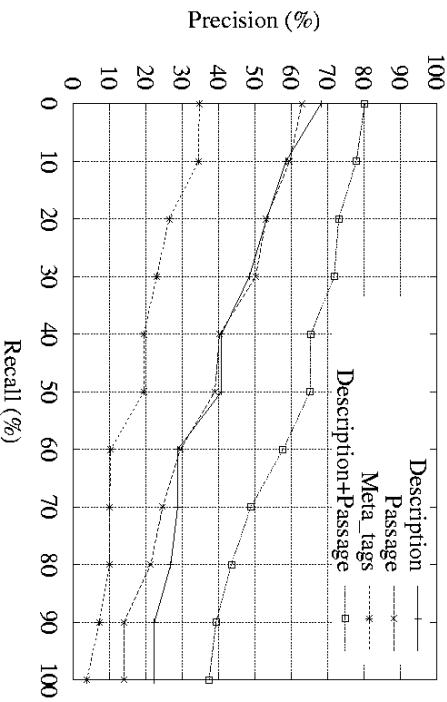
P. Calado, E. Moura, B. Ribeiro-Neto, I. Silva, N. Ziviani. Local versus Global Information in the Web. *ACM SIGIR 2000 e ACM TOIS 2003*

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

9

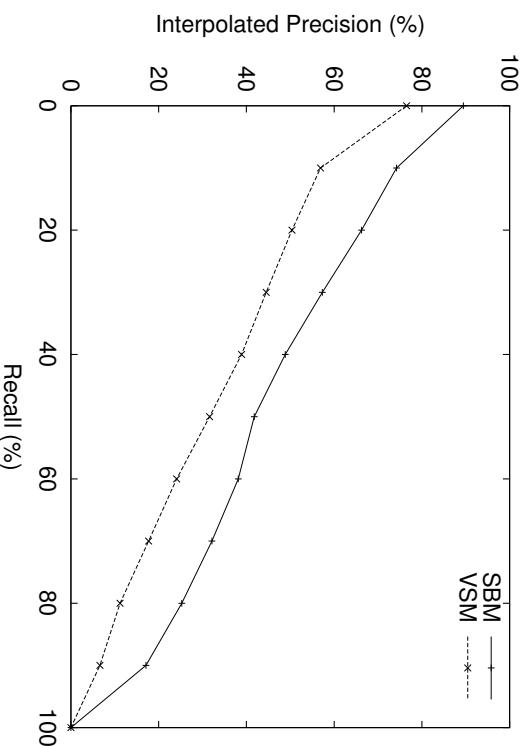
Image Retrieval

Average precision for 25 queries



Coelho, T., Calado, P., Souza, L., Ribeiro-Neto, B., and Muntz, R. Image retrieval using multiple evidence ranking. *IEEE TKDE*, 2003

Modelos de RI



Pôssas, Ziviani, Meira Jr. e Ribeiro-Neto *Set-Based Model: A New Approach for Information Retrieval, ACM SIGIR 2002 e ACM TOIS 2005*

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

11

Web Classification

Link Similarity Measures	Cade12			Cade188		
	KNN	SVM	NB	KNN	SVM	NB
B. Coupling	36.31	40.02	39.30	22.32	23.08	22.70
Amsler	81.26	77.65	58.80	70.57	68.91	47.01
Co-citation	81.55	77.89	59.03	71.07	69.53	47.31
Companion	73.00	63.66	42.76	68.54	73.63	29.82
Baseline	39.45	40.86	39.38	24.45	24.31	22.82

Calado, Cristo, Moura, Gonçalves, Ziviani, and Ribeiro-Neto, *Linkage Similarity Measures for the Classification of Web Documents, JASIST 57(2), 2006, 208-221.*

Coluto, Cristo, Gonçalves, Calado, Ziviani, Moura, Ribeiro-Neto *A Comparative Study of Citations and Links in Document Classification, ACM/IEEE JCDL 2006, 75-84.*

Content-Targeted Advertising

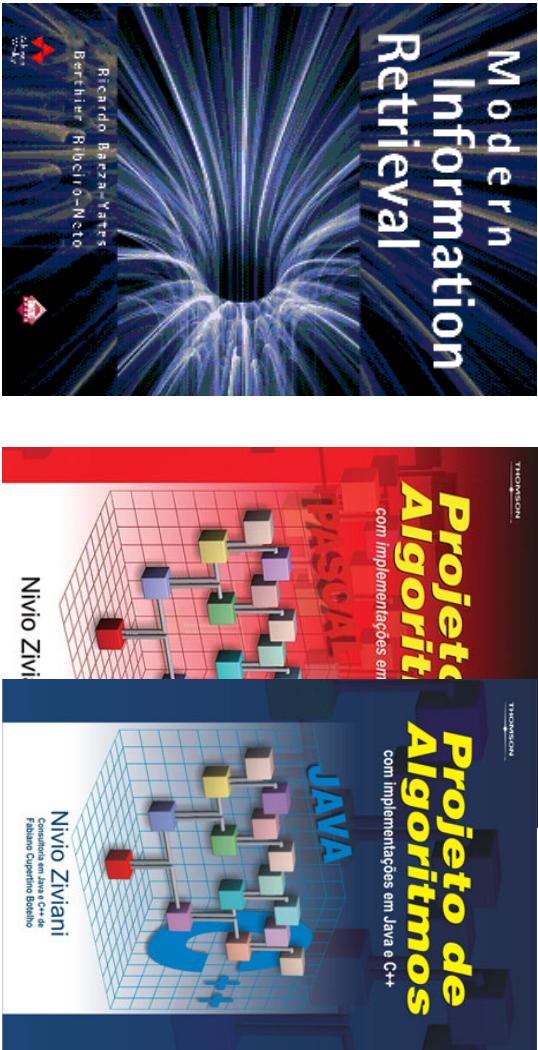
A Genetic Programming Approach

Methods	pavg@3
Score	Gain
AAK_H (Baseline)	0.314
GP	0.508 +61.7%

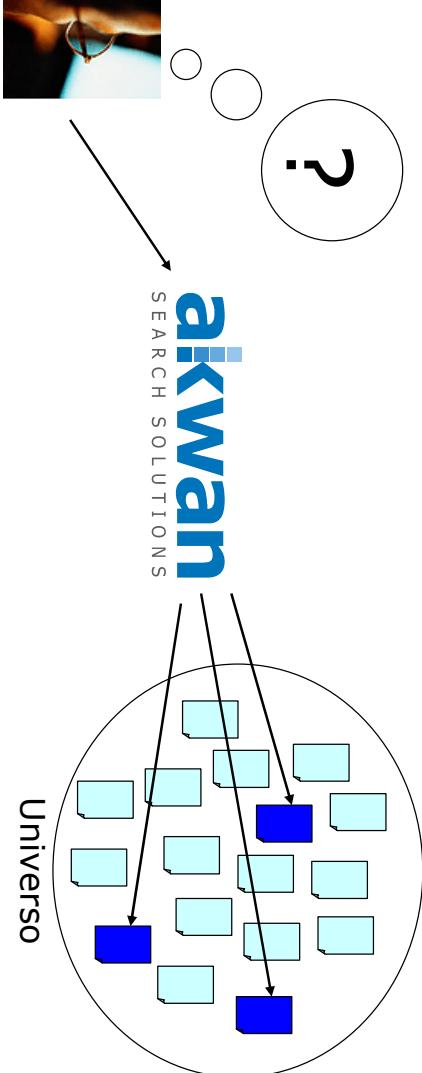
Lacerda, Cristo, Gonçalves, Fan, Ziviani e Ribeiro-Neto. Learning to Advertise, ACM SIGIR 2006, Agosto, Seattle, USA, 549-556.

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG) 13

DCC/UFMG: Livros Publicados



Spin offs do DCC-UFMG



www.akwan.com.br

Miner - Aplicativos para a Web - Netscape

Smart Price - Microsoft Internet Explorer

Metasearch Website - Microsoft Internet Explorer

Akwan Information Technologies - Microsoft Internet Explorer

Address <http://www.akwan.com.br/>

Home Services Products Clientes Inquiries Contact

NOTA DIRETA DE LEIAV: Go Links

akwan
INFORMATION TECHNOLOGIES

A Akwan desenvolve e implementa soluções de acesso à informação sob medida.

As soluções são de dois tipos principais:

- **maquinhas de busca na Web**, incluindo busca em todo um país;
- **sistemas de organização da informação** de negócios.

As soluções Akwan são adaptadas para as necessidades de cada cliente, o que permite ao usuário localizar informação de seu interesse com rapidez e presteza.

Nossas soluções estão presentes em instituições como Banco Real, Bovespa, FIEP, UOL, IBEST, IG, Uol, no Brasil, e Banco Bilbao Vizcaya Argentaria, La Caixa, na Espanha.

ENGLISH VERSION WEBMASTER

© 2001 Akwan Information Technologies

Organização de Informação
Organize de forma imediatamente automaticamente sua informação de negócios. A base de informação é composta por documentos corporativos e documentos diversos como PDF, HTML, Word, etc.

Maquinhas de Busca
Maquinhas de busca, incluindo busca local e vertical, para toda a web de um país ou de empresas sob medida para suportar o seu negócio.

Internet

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

E a akwan virou Google ...

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

17

www.akwan.com.br



Akwan é agora parte do Google Brasil.
[Saiba mais.](#)

- **Pesquisa de web específica ao país:** Se você está interessado em buscadores que são específicos a um país, tente [Google Ferramentas de idiomas](#) para visitar o site do Google no seu domínio local, como o [Google Brasil](#).
- **Pesquisa de web no seu site:** Se você está interessado em providenciar pesquisa de web no seu site, aprenda sobre [Google AdSense para pesquisas](#).

 ENGLISH VERSION

©2005 Google



LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

21

O Que Disse a Mídia

Negócio premia esforço de pesquisa no Brasil
(Heloisa Magalhães e Ricardo Cesar, Valor Econômico, 21/07/2005)

A venda da Akwan pode oxigenar o mercado de capital de risco no Brasil (IstoÉ Dinheiro, 27/07/2005)

Alguns Clientes Akwan



LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

23

As Trajetórias da Akwan e da Google

Ambas as empresas:

- Tiveram origem em universidades de reconhecida excelência acadêmica em seus respectivos países
- Receberam nomes com significados emblemáticos
- Utilizavam uma estrutura operacional baseada em *clusters* de computadores
- Receberam aporte de capital de risco para início de suas atividades
- Superaram a bolha da Internet gerando receita a partir de sua própria tecnologia
- Detinham parte significativa de seus respectivos mercados

“Não foi somente uma questão de escolher o Brasil. Escolhemos a Akwan que por acaso fica no país.”
(Alan Eustace, VP de Engenharia da Google, Exame, 03/08/2005)

Não se Tratam de Casos Isolados

Universidades de ponta nos EUA:

- Yahoo: criada por David Filo e Jerry Yang – Stanford 1994
- Inktomi: criada pelo Prof. Erik Brewer, Berkeley
- MetaCrawler: criada por Erik Selberg (aluno) e Oren Etzione (orientador), Univ. Washington 1995
- Netbot: criada pelo Prof. Dan Welch, Univ. Washington 1996

Revista Fortune (1999):

- MIT: 4.000 empresas, US\$300 bilhões
- Stanford: *spin-offs* no valor de US\$1,1 trilhão

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

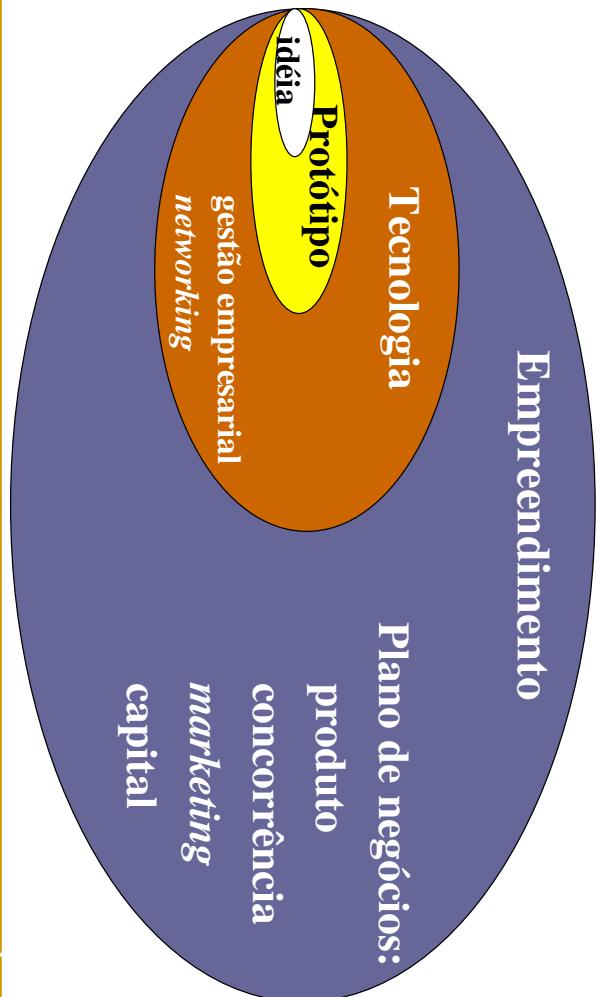
25

Modelo na UFMG / DCC: A Criação de Empresas Start-ups

- não é trivial
- envolve recursos e mecanismos operacionais extra-muros
- gestão empresarial
- *marketing*
- *networking*
 - **acesso** a altos círculos de negócios
 - reconhecimento pelo mercado
 - facilidade para atrair capital de risco
- no mundo da Internet e da Web:
 - desafio amplificado pela velocidade com que idéias são transformadas em produtos e serviços

Criando empreendimentos

Modelo UFMG / DCC: A Criação de Empresas *Start-ups*



LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

27

Transformação de uma *idéia* em uma *start-up*

Idéia → **Protótipo** → **Tecnologia** → **Start-up**

- **Idéia:** resultado de uma pesquisa
- **Protótipo:** prova de conceito em pré-fase de alpha teste
- **Tecnologia:** inclui todo um “empacotamento”
- **Start-up:**
 - Aporte de *capital externo*
 - *Plano de negócios:* o modelo de negócios, análise da concorrência, marketing, plano de gerência do negócio que seja escalável, cronograma físico financeiro, apropriação de custos, plano de contratação e gerência de pessoal, advocacia, etc.

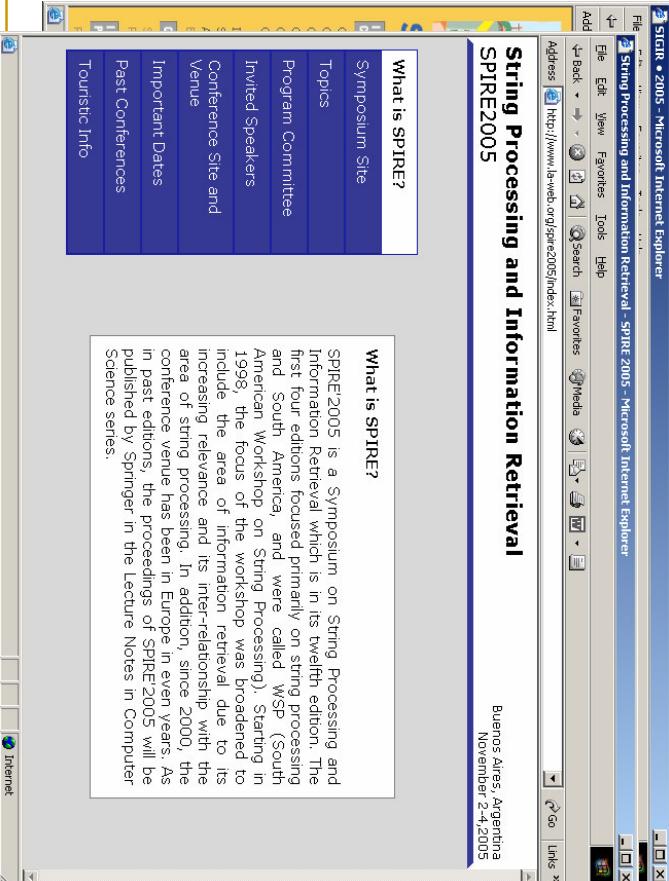
Participação da Universidade

- Órgãos da UFMG podem participar na *tecnologia e no empreendimento*
 - Start-up é Sociedade Anônima
→ Órgãos da UFMG não são responsáveis por eventuais prejuízos e responsabilidades financeiras
 - Transferência de ações da start-up para UFMG
 - Além de formadora de recursos humanos, é fonte de empreendimentos e geradora de riqueza

DCC/UFMG: Disciplinas de RI e Correlatas

- Recuperação de Informação A
 - Berthier Ribeiro-Neto
- Recuperação de Informação B
 - Nívio Ziviani
- Gerência de Dados na Web
 - Alberto Laender
 - Mineração de Dados
 - Wagner Meira Jr.
- Bibliotecas Digitais
 - Marcos Gonçalves

DCC/UFMG: Conferências Organizadas RI



LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

31

Participação Societária

- Sócios fundadores (empreendedores)
 - Alberto Laender, Berthier Ribeiro-Neto, Ivan Moura Campos, Nívio Ziviani (DCC/UFMG)
 - Guilherme Emrich, Marcus Regueira (FIR Capital Partners)
- Sócios participantes
 - Ex-alunos da UFMG
- Sócio institucional
 - UFMG (FUNDEP)

O Investidor de Risco

- Guilherme Emrich e Marcus Regueira - *angels*
 - ▣ Andy Bechtolsheim: *angel* na Google (US\$100 mil)
- Além de aportar recursos:
 - ▣ Fizeram o *coaching* dos empreendedores
 - ▣ Ajudaram na busca de clientes
 - ▣ Negociaram a venda da empresa
- É possível gerar retorno com *venture capital* no país

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

33

Modelo de Atuação da Akwan

1. Associação com uma universidade de comprovada excelência acadêmica
2. Cooperação com um grupo de pesquisa reconhecido internacionalmente
3. Seleção dos melhores estudantes para estágio
4. Contratação dos melhores estagiários

“Nosso único critério é encontrar as melhores pessoas.”
(Alan Eustace, VP de Engenharia da Google, Exame, 03/08/2005)

Lições Aprendidas

- Mais de 280 incubadoras (maior parte com fortes ligações com universidades)
- Finep, BNDES, FAPs envolvidos na criação e fomento de empresas inovadoras
- Petros: R\$110 milhões para 5 fundos de investimento
- Petrobrás – programa de P&D por meio do Cenpes-Universidades (por exemplo COPPE/UFRJ)
- Porto Digital em Recife

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

35

Referências

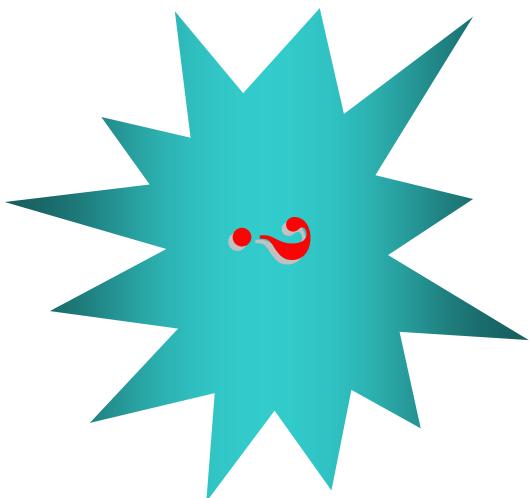
- J.A. Deutscher, T. Renault e N. Ziviani, A Geração de Riqueza a Partir da Universidade: O Caso Akwan. *Inteligência Empresarial* 24 (3), Jul/Ago/Set 2005, Páginas 2-8, COPPE/UFRJ, Rio de Janeiro.
- N. Ziviani, As Origens da Família Miner. *Boletim da UFMG*, Número 1270, Ano 26, 29/03/2000, Página 2.
- Jeremy Kahn, At Cambridge A New Age. *Fortune*, March 1999, 60-64.

Obs.: duas primeiras referências podem ser obtidas em

www.dcc.ufmg.br/~nivio

Conclusões

1. Gerar tecnologia de ponta exige parceria com um grupo de pesquisa de ponta
2. Os *angels* exercem papel muito importante no surgimento de empresas intensivas em conhecimento
3. Atuação do Governo: falta promover mudanças no ambiente institucional que favoreça a inovação e o *funding* dos empreendimentos
4. O exemplo da Akwan aponta um caminho alternativo para fortalecer o sistema de C&T no país
5. O modelo da hélice tripla deve fazer parte das políticas industriais de C&T para o país



Centro de P&D Google Brasil

- Engenheiros têm que ter pelo menos Graduação
 - Computação, Engenharia Elétrica, Física
- Podem estar cursando Mestrado ou Doutorado
- Processo de seleção pode levar de 8-16 semanas
- Inscreva-se já!

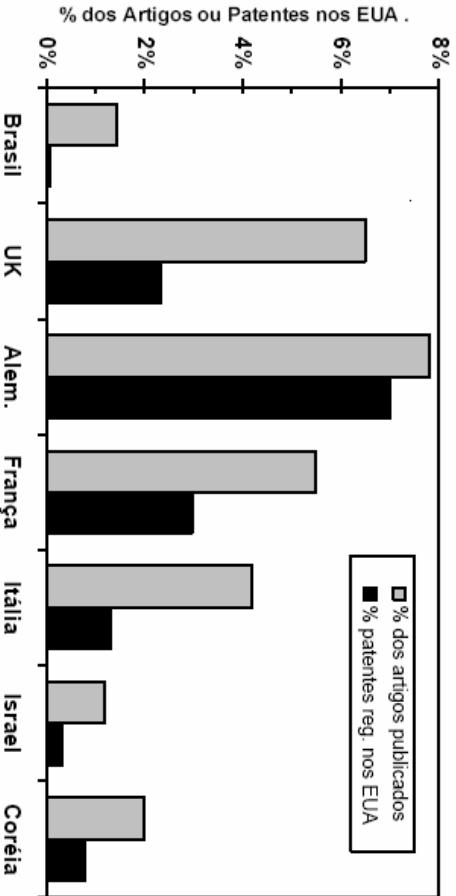
<http://www.google.com.br/jobs>
Software Engineer

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

39

Cenário Mundial de P&D

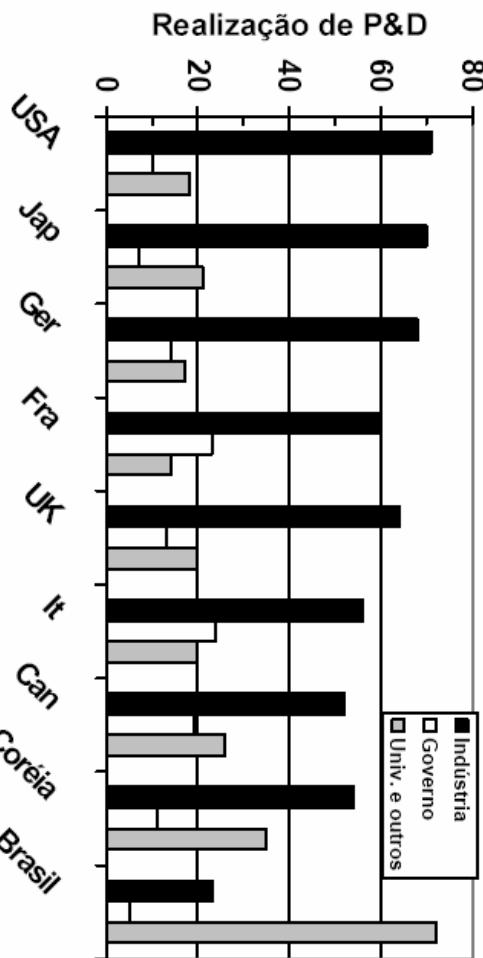
Artigos X Patentes



Fonte: A Universidade, a Pesquisa e a Empresa
(Brito Cruz, 2003)

Cenário Mundial de P&D

Esfórcio Realizado



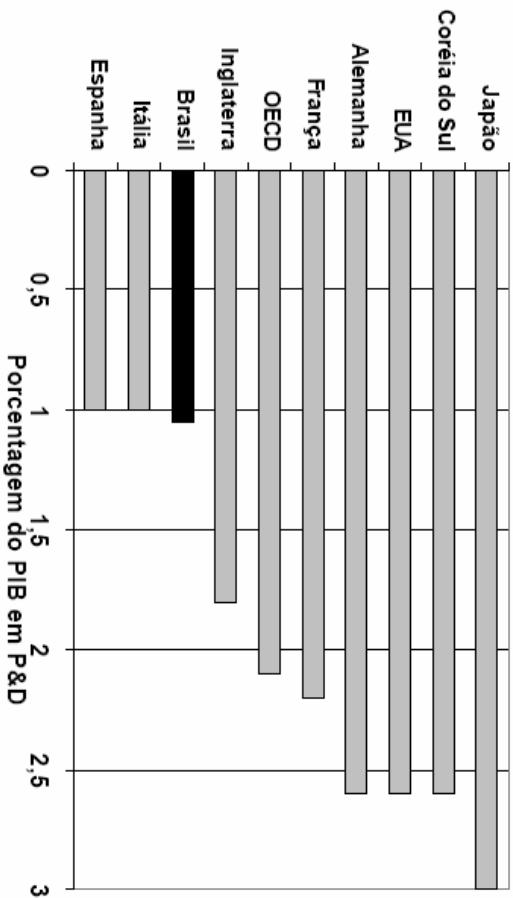
Fonte: A Universidade, a Empresa e a Pesquisa
(Brito Cruz, 2003)

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

41

Cenário Mundial de P&D

Investimento



Fonte: A Universidade, a Pesquisa e a Empresa
(Brito Cruz, 2003)

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

42

Por que empreender?

O Plano Empresarial

- Empresas de sucesso se engajam, desde cedo, em um processo de **criação de valor**
- **Valor** materializa-se, mais cedo ou mais tarde, em:
 - um grande ambiente de trabalho
 - um trabalho que propicia satisfação pessoal e reconhecimento profissional
 - recompensa financeira
 - empregos de melhor qualidade e, neste sentido, mais nobres

Breve Histórico

- Fevereiro de 1998 – Victor Ribeiro defende dissertação
- Abril de 1998 – Criação da Miner
- Junho/1999 – Venda da Miner para o grupo Folha/UOL
- Novembro/1999 – Lançamento do TodoBR e do TodoCL
- Março/2000 – Lançamento do Buscopio na Espanha
- Maio/2000 – Criação da Akwan
- Julho/2005 – Venda da Akwan para a Google

Spin-offs do DCC

- Miner Technology Group (1998 – 1999)
(www.miner.com.br)
- Akwan Information Technologies (2000 – 2005)
(www.akwan.com.br)
- MetaSys (2003 -)
(www.metasys.com.br)
- Smart Price (2002 -)
(www.smartprice.com.br)

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

45

Exemplos de Tópicos de Pesquisa

- **Modelos de RI:** determinam a precisão da resposta a consultas e formam a base tecnológica do componente principal de qualquer sistema de RI, o processador de consultas
- **Eficiência em RI:** Sistemas de RI, além de eficazes, devem ser extremamente eficientes, uma vez que o tempo de resposta é um problema central.

DCC/UFMG: Conferências Organizadas RI

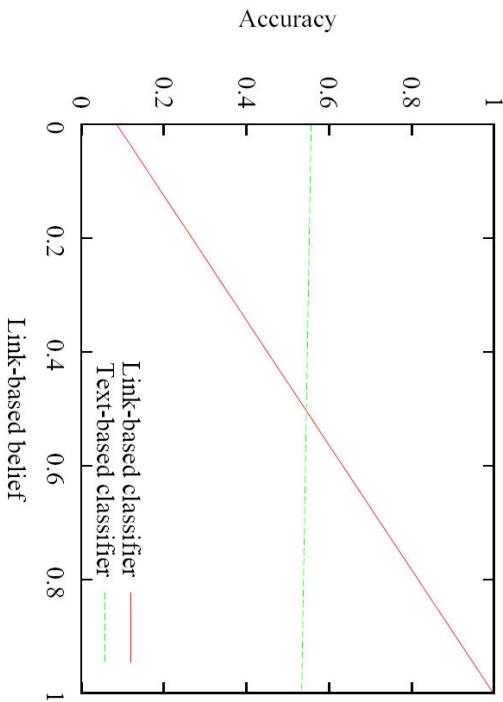
- 28th ACM SIGIR on Research and Development in Information Retrieval – Salvador, Brasil (www.sigir2005.org)

- SPIRE – String Processing and Information Retrieval – Buenos Aires, Argentina (www.la-web.org/spire2005/)

LATIN – Laboratório para Tratamento da Informação (DCC/UFMG)

47

Classificação de documentos utilizando links derivados de citações e hiperlinks



Couto, Cristo, Gonçalves, Calado, Ziviani, Moura, Ribeiro-Neto A Comparative Study of Citations and Links in Document Classification, *ACM/IEEE-CS JCDL 2006*, 75-84.