



RI na Web

Prof. Luciano Barbosa

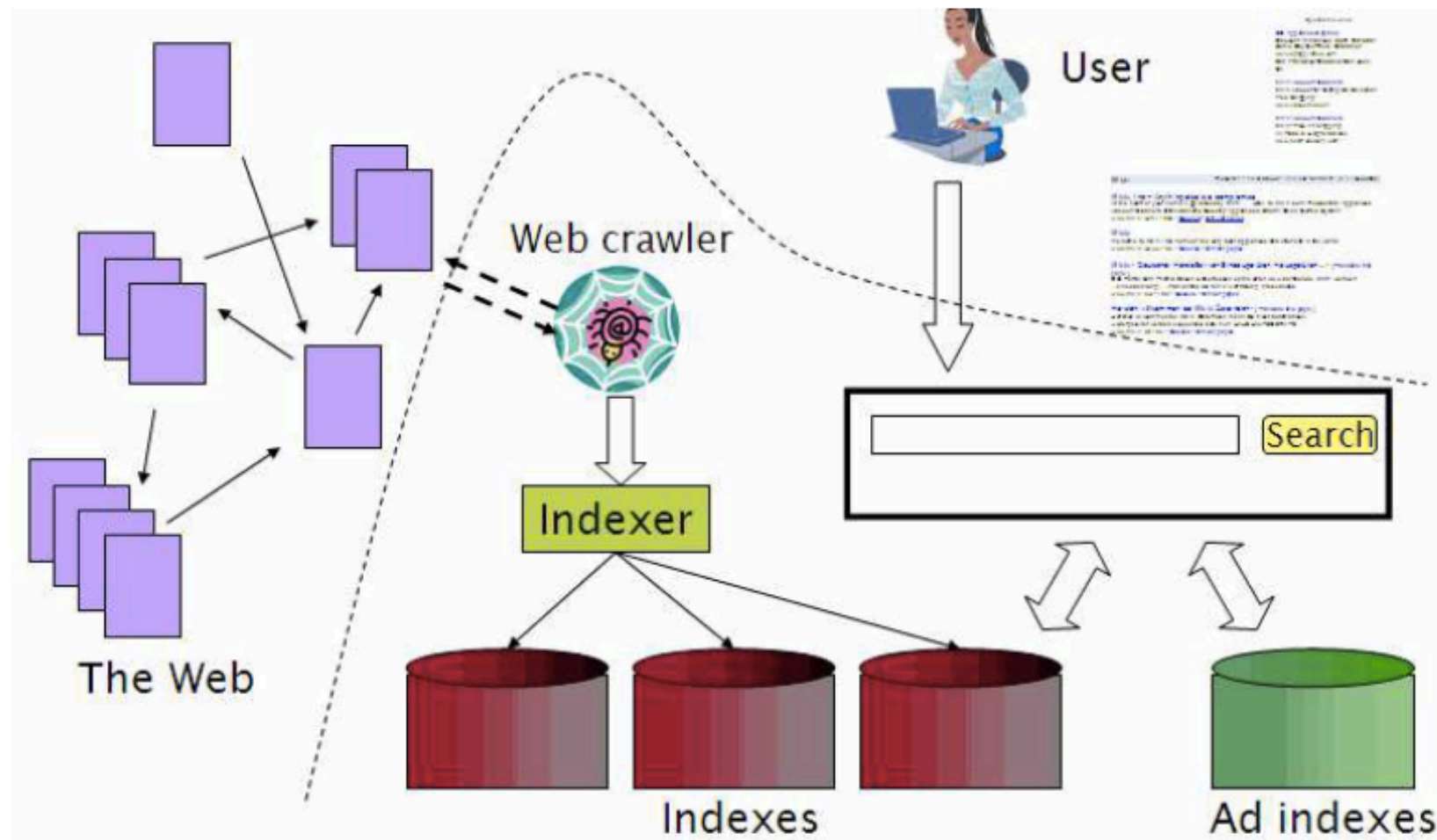
(Parte do material retirado dos slides dos livros adotados)



UNIVERSIDADE
FEDERAL
DE PERNAMBUCO



Visão Geral





Anúncios em Engenhos de Busca

- Objetivo: encontrar o melhor anúncio para um dado usuário e contexto
- Muito lucrativo pra engenhos de busca e lojas online
- Intersecção de áreas como RI, ML, otimização e micro-economia



Tipos de Anúncios

- Busca patrocinada: resposta a uma busca a um engenho de busca
- Casamento de conteúdo: coloca anúncios em páginas
- Banners



Primeira Geração de Busca Patrocinada: Goto (1996)



- Páginas ranqueadas de acordo com o lance -> maximizar o retorno para Goto
- Nenhuma separação entre anúncios e resultados
- Sem nenhum ranqueamento de relevância
- Ex: Buddy Blake paga um máximo de \$0.38 quando uma pessoa clica no link



Segunda Geração: Google AdWords (2000/2001)

- Separação entre os resultados e os anúncios

Web Images Maps News Shopping Gmail more Sign in

Google discount broker Search Advanced Search Preferences

Web Results 1 - 10 of about 807,000 for discount broker [definition]. (0.12 seconds)

Discount Broker Reviews
Information on online discount brokers emphasizing rates, charges, and customer comments and complaints.
www.broker-reviews.us/ - 94k - Cached - Similar pages

Discount Broker Rankings (2008 Broker Survey) at SmartMoney.com
Discount Brokers. Rank/ Brokerage/ Minimum to Open Account, Comments, Standard Commission*, Reduced Commission, Account Fee Per Year (How to Avoid), Avg. ...
www.smartmoney.com/brokers/index.cfm?story=2004-discount-table - 121k - Cached - Similar pages

Stock Brokers | Discount Brokers | Online Brokers
Most Recommended. Top 5 Brokers headlines. 10. Don't Pay Your Broker for Free Funds May 15 at 3:39 PM. 5. Don't Discount the Discounters Apr 18 at 2:41 PM ...
www.fool.com/investing/brokers/index.aspx - 44k - Cached - Similar pages

Discount Broker
Discount Broker - Definition of Discount Broker on Investopedia - A stockbroker who carries out buy and sell orders at a reduced commission compared to a ...
www.investopedia.com/terms/d/discountbroker.asp - 31k - Cached - Similar pages

Discount Brokerage and Online Trading for Smart Stock Market ...
Online stock broker SogoTrade offers the best in discount brokerage investing. Get stock market quotes from this internet stock trading company.
www.sogotrade.com/ - 39k - Cached - Similar pages

15 questions to ask discount brokers - MSN Money
Jan 11, 2004 ... If you're not big on hand-holding when it comes to investing, a discount broker can be an economical way to go. Just be sure to ask these ...
moneycentral.msn.com/content/Investing/StartInvesting/P66171.asp - 34k - Cached - Similar pages

Sponsored Links

Rated #1 Online Broker
No Minimums. No Inactivity Fee
Transfer to Firsttrade for Free!
www.firsttrade.com

Discount Broker
Commission free trades for 30 days.
No maintenance fees. Sign up now.
TDAMERITRADE.com

TradeKing - Online Broker
\$4.95 per Trade, Market or Limit
SmartMoney Top Discount Broker 2007
www.TradeKing.com

Scottrade Brokerage
\$7 Trades, No Share Limit. In-Depth Research. Start Trading Online Now!
www.Scottrade.com

Stock trades \$1.00 to \$3
100 free trades, up to \$100 back for transfer costs. \$500 minimum
www.sogotrade.com

\$3.95 Online Stock Trades
Market/Limit Orders, No Share Limit and No Inactivity Fees
www.Marsco.com

INGDIRECT | ShareBuilder

Sogo Trade aparece no conjunto de resultados

Sogo Trade aparece no conjunto de anúncios

Engenhos de busca dizem não dar prioridade no ranking de resultados para anúncios



Where's Google making its money?

\$32.2 billion
in total advertising revenue

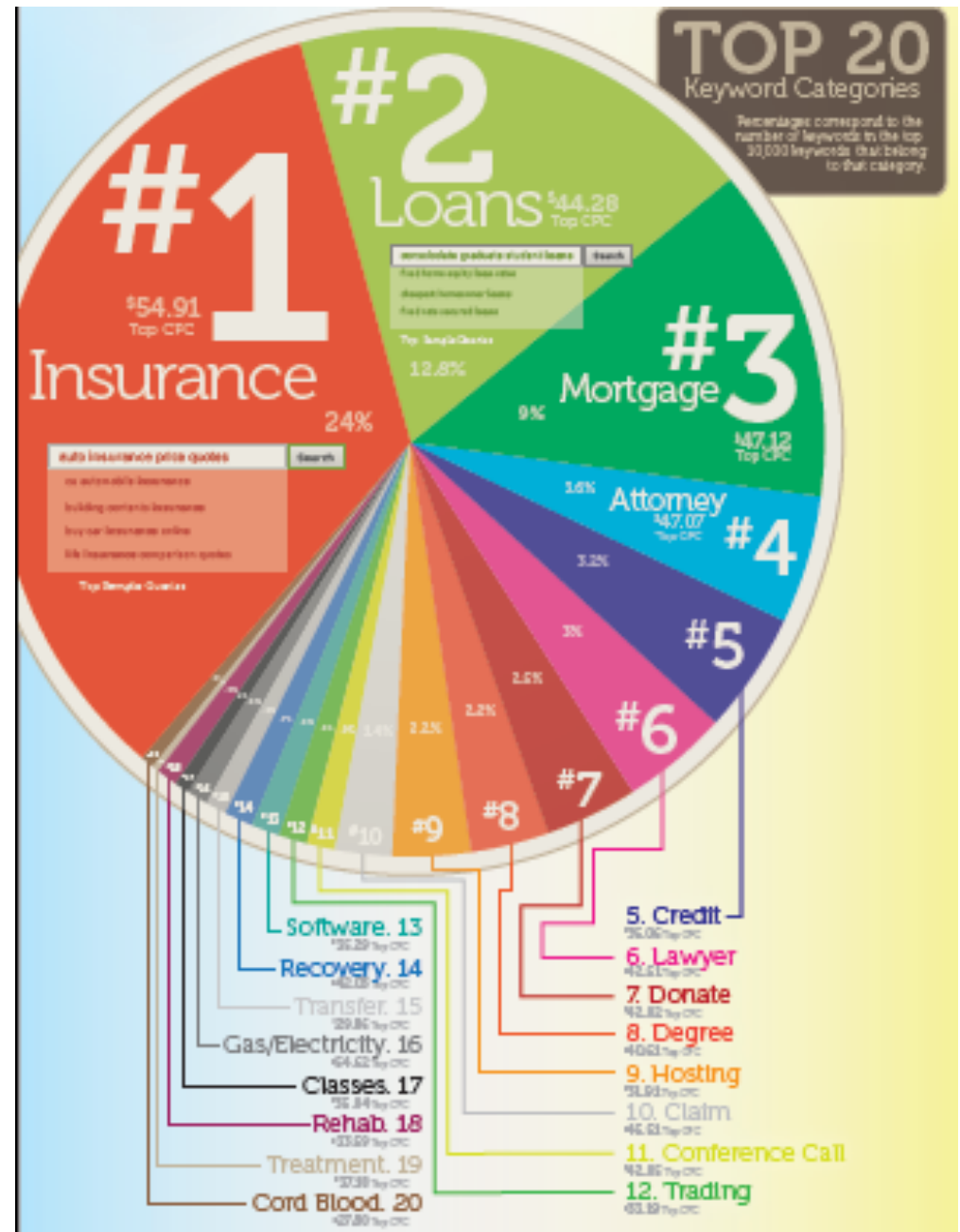
SPOILER ALERT: **ADVERTISING!**

\$33.3 billion
in total revenue in Q3 2013 - Q2 2014

of Google's
Revenue

97%
is from
advertising

Top 20
Most Expensive
Keywords
in Google AdWords Advertising





Como Anúncios são Ranqueados

- Anunciantes dão um lance por palavras-chave: venda por leilão
- Qualquer um pode participar do leilão
- Anunciantes só são cobrados se o usuário clicar no seu anúncio
- Estratégia simples de ranqueamento
 - Baseado no valor do lance (similar a Goto)
 - Pode haver abuso
 - Ex: consulta: “tratamento do câncer” -> anúncio: “como escrever seu testamento”
 - Evitar mostrar anúncios não relevantes ou ofensivos



Como Anúncios são Ranqueados

- Ranquear baseado no preço do lance e relevância
- Medida de relevância: taxa de cliques (clickthrough rate - CTR)
 - Um anúncio não relevante vai ser ranqueado baixo
 - Pode diminuir o retorno a curto prazo pro EB
 - Maior retorno a longo prazo (maior volume de cliques)
- Outros fatores
 - Localização, hora do dia, qualidade e tempo de carregamento da página final



Google's second price auction

- Funcionamento
 - Cada anunciante dá um lance
 - Quem deu o maior lance leva
 - Paga o lance do segundo colocado
- Induz ao preço justo



Google's second price auction

- Lance igual ao valor justo: $l_i = v$ (estratégia ótima)
 - Se alguém fizer menor lance, vc paga esse menor lance
 - Se alguém fizer o mesmo lance, nenhum benefício
- Lance maior que v : $l_i > v$
 - Aumenta a chance de vencer mas o valor a ser pago depende do segundo colocado
 - Alguém fizer lance maior que v e menor que o seu: $l_i > l_j > v$: vc ganha mas vc vai pagar mais do que o justo
- Lance menor que v :
 - Diminui a chance de vencer



Google's second price auction

advertiser	bid	CTR	ad rank	rank	paid
A	\$4.00	0.01	0.04	4	(minimum)
B	\$3.00	0.03	0.09	2	\$2.68
C	\$2.00	0.06	0.12	1	\$1.51
D	\$1.00	0.08	0.08	3	\$0.51

- Bid: lance máximo por um clique
- CTR: porcentagem de vezes que usuários clicam no anúncio (medida de relevância)
- ad rank: bid X CTR
 - Quanto o anunciante está disposto a pagar vs a relevância do anúncio
- Rank
- Paid: second price auction pago pelo anunciante



Google's second price auction

advertiser	bid	CTR	ad rank	rank	paid
A	\$4.00	0.01	0.04	4	(minimum)
B	\$3.00	0.03	0.09	2	\$2.68
C	\$2.00	0.06	0.12	1	\$1.51
D	\$1.00	0.08	0.08	3	\$0.51

- Second price auction
 - Quantidade mínima necessária para manter a posição (mais 1 centavo)
 - Exemplo para anunciante D
 - ad rank do anunciante anterior: 0.04 (A)
 - CTR: 0.08
 - Valor pago X CTR = ad rank -> Valor pago X 0.08 = 0.04 -> Valor pago = $0.5 + 0.01 = 0.51$



Anúncios: Pros e Cons

- Pros
 - Engenho de busca obtém receita toda vez que o usuário clica no anúncio
 - Anúncios são apresentados baseado em relevância: pune anúncios ruins
 - Anunciantes encontram novos usuários
- Cons
 - Keyword arbitrage (spam)
 - Compra uma palavra-chave
 - Redireciona para um sites com anúncios



Problemas com Anúncios

- Violação de marcas registradas
 - Ex: Geico
 - Durante parte de 2005 o termo geico no Google foi comprado pelos competidores
 - Geico processou mas perdeu nos EUA
 - Caso similar com Luis Vuitton na Europa



Consultas em Engenhos de Busca

Most frequent queries on a large search engine on 2002.10.26.

1	sex	16	crack	31	juegos	46	Caramail
2	(artifact)	17	games	32	nude	47	msn
3	(artifact)	18	pussy	33	music	48	jennifer lopez
4	porno	19	cracks	34	musica	49	tits
5	mp3	20	lolita	35	anal	50	free porn
6	Halloween	21	britney spears	36	free6	51	cheats
7	sexo	22	ebay	37	avril lavigne	52	yahoo.com
8	chat	23	sexe	38	hotmail.com	53	eminem
9	porn	24	Pamela Anderson	39	winzip	54	Christina Aguilera
10	yahoo	25	warez	40	fuck	55	incest
11	KaZaA	26	divx	41	wallpaper	56	letras de canciones
12	xxx	27	gay	42	hotmail.com	57	hardcore
13	Hentai	28	harry potter	43	postales	58	weather
14	lyrics	29	playboy	44	shakira	59	wallpapers
15	hotmail	30	lolas	45	traductor	60	lingerie



Distribuição das Consultas

- Distribuição power law:
 - Poucas consultas frequentes e muitas consultas pouco frequentes
 - Útil para cache de consultas
- Exemplos de consultas raras:
 - Busca por nomes de pessoas, cidades, livros etc



Tipos de Consultas

- **Informacional:**
 - Consulta sobre um tópico
 - Pode estar presente em várias páginas
 - Ex: Informação sobre doenças
- **Navegacional**
 - Encontrar uma página específica
 - Ex: Receita Federal, United Airlines
- **Transacional:**
 - Encontrar sites para comprar ou baixar algo
 - Ex: comprar xbox
 - Ex: baixar acrobat reader
- **Engenho de busca precisa identificar o tipo de consulta**



Perfil dos Usuários

- Consultas curtas (média < 3)
- Raramente usam operadores
- Não querem gastar muito tempo escrevendo consultas
- Dão preferência aos primeiros resultados
- Querem uma interface simples

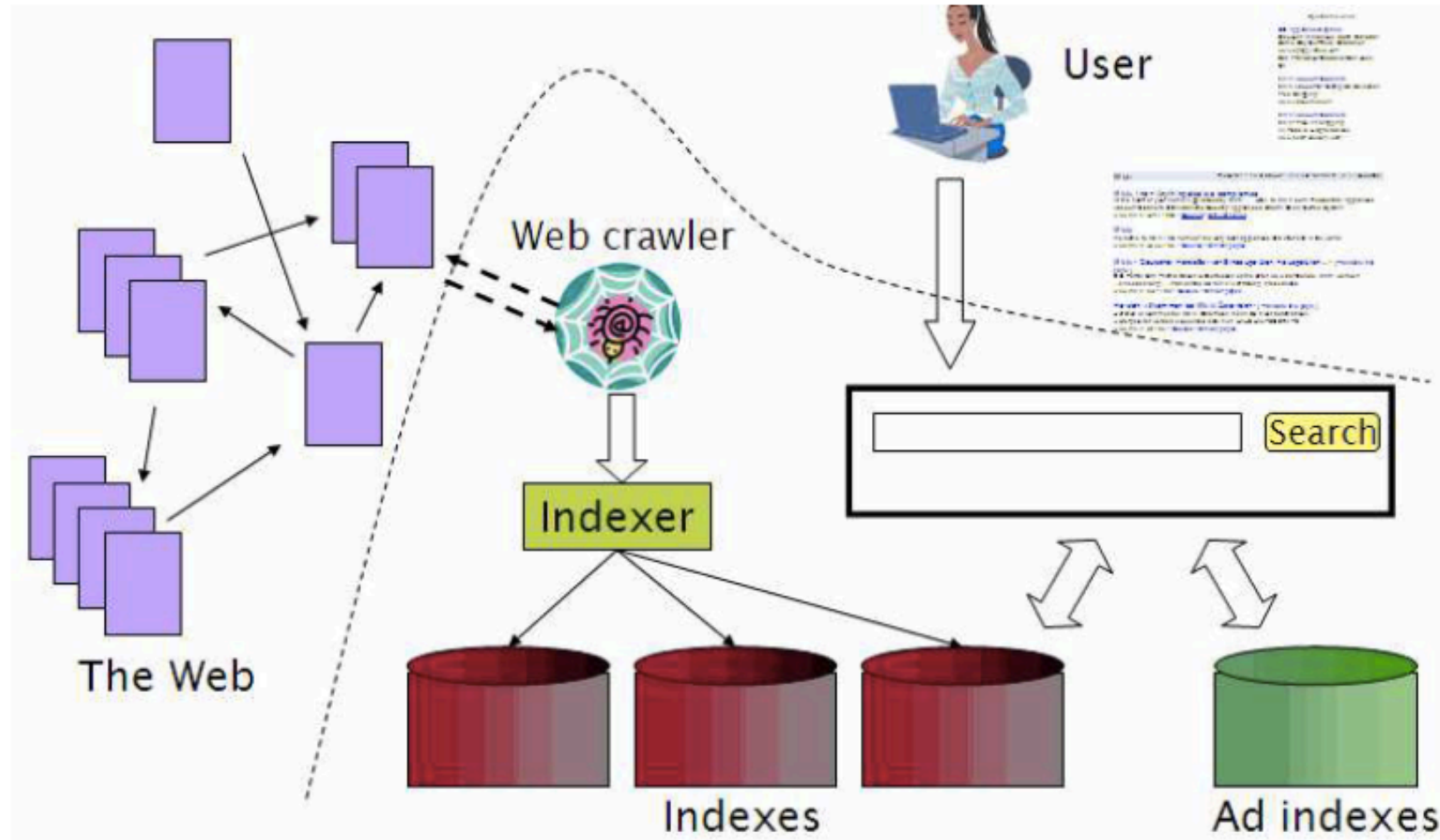


Como o Usuário Avalia Engenhos de Busca?

- Precisão
- Para algumas consultas recall é importante
 - Ex: busca por patentes, por emprego etc
- Tempo de resposta
- Sem pop-ups



O que Vimos: Como Engenheiros de Busca Funcionam





O que Vimos

- Coleta de páginas na Web
- Processar texto: tokenização, encoding etc
- Classificar texto
- Extrair estrutura de páginas HTML
- Arquivos invertidos
- Modelos de RI: similaridade entre textos



O que Vimos

- Aprender a ranquear
- Refinar consultas:
 - Similaridade de strings
 - Language model
 - Mutual information
 - Relevance feedback
- Avaliar ranking: diferentes métricas
- Busca social e sistemas de recomendação
- Autoridade em grafos: PageRank e HITs
- Anúncios em EB

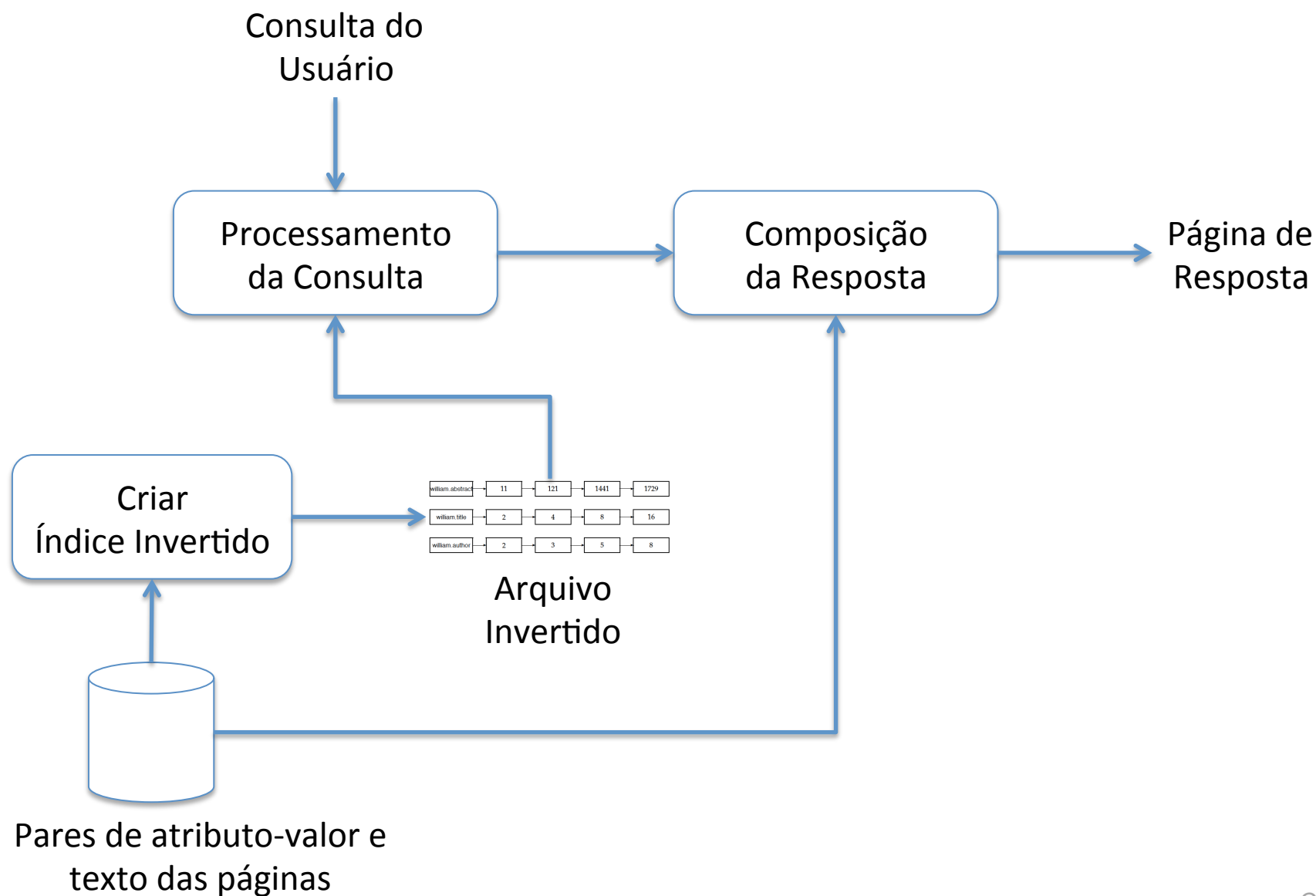


Em Resumo: Diversas Áreas

- Mineração na Web
- Mineração de Texto
- Estrutura de dados
- Grafos



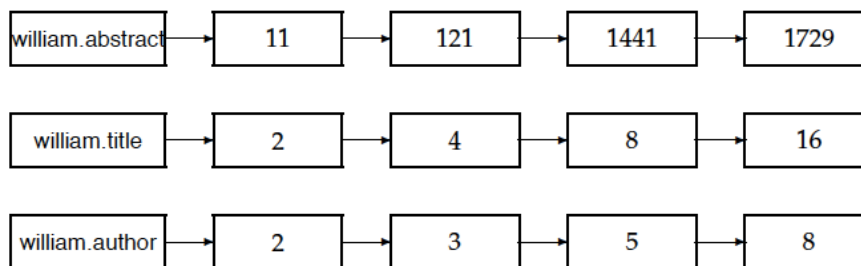
Projeto: Parte 2





Criar Arquivo Invertido

- Pré-processamento:
 - Pares atributo-valor: 5 atributos mais frequentes
 - Mapear atributos entre sites: ex.: quartos -> qts
 - Discretizar atributos numéricos: quartis [0-25],(25-50],[50-75],[75-100]
- Índice invertido
 - Frequência
 - Duas partes:
 - Field index: nome do campo adicionado ao termo
 - Termos do documento



- Duas versões: sem compressão e com compressão com intervalo (medir tamanho)

COMPUTER: 283154, 283159, 283202, ➡ COMPUTER: 283154, 5, 43



Processamento de Consulta

- Entrada: consulta do usuário
- Compor consulta: campo+palavras ou só palavras
- Ler postings: escolher document-at-a-time ou term-at-a-time
- Ranquear documentos usando Modelo de Espaço de Vetores
 - Peso: com e sem tfidf
- Implementar correlação de spearman ou kendal tau para comparar o ranking de 5 consultas



Composição da Resposta

- Entrada: ranking
- Ler informações dos documentos a serem apresentadas:
 - URL e pares de atributo-valor
- Criar interface: busca nos atributos e no texto
- Recomendar para cada atributo as 3 palavras com maior mutual information

Bibliographic Search

Search category	Value
Author	Example: Widom, J or Garcia-Molina <input type="text"/>
Title	Also a part of the title possible <input type="text"/>
Date of publication	Example: 1997 or <1997 or >1997 limits the search to the documents appeared in, before and after 1997 respectively <input type="text"/>
Language	Language the document was written in English <input type="button" value="v"/>
Project	ANY <input type="button" value="v"/>
Type	ANY <input type="button" value="v"/>
Subject group	ANY <input type="button" value="v"/>
Sorted by	Date of publication <input type="button" value="v"/>