



Recuperação de Informação

Edleno Silva de Moura

O curso:

- Base teórica sobre a área (Modelos, Avaliação, Algoritmos e Estruturas de dados Importantes)
- Sistemas de **BUSCA**
- **Bibliotecas Digitais**
- **Busca na Web**



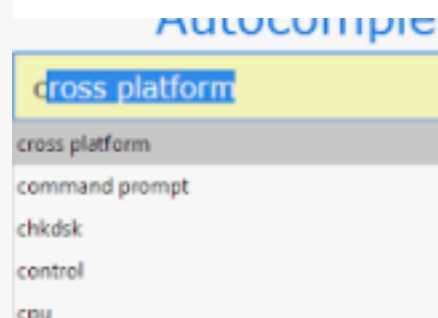
FOCO EM BUSCA



Busca em todos os lugares



search as you
search as you type
search as you type elasticsearch
search as you type javascript
search as you type jquery
search as you type angularjs



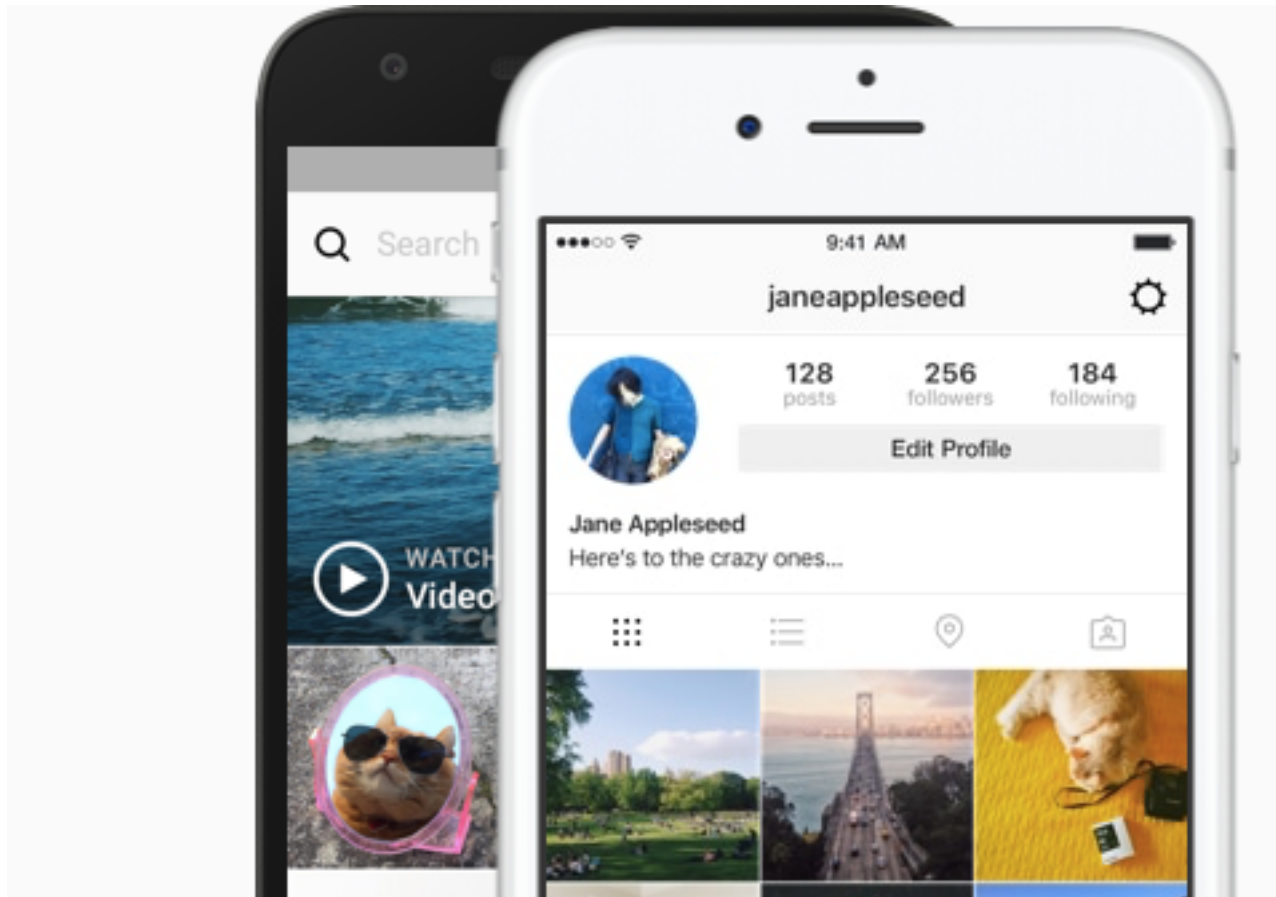
Busca (linkedin, facebook, americanas)



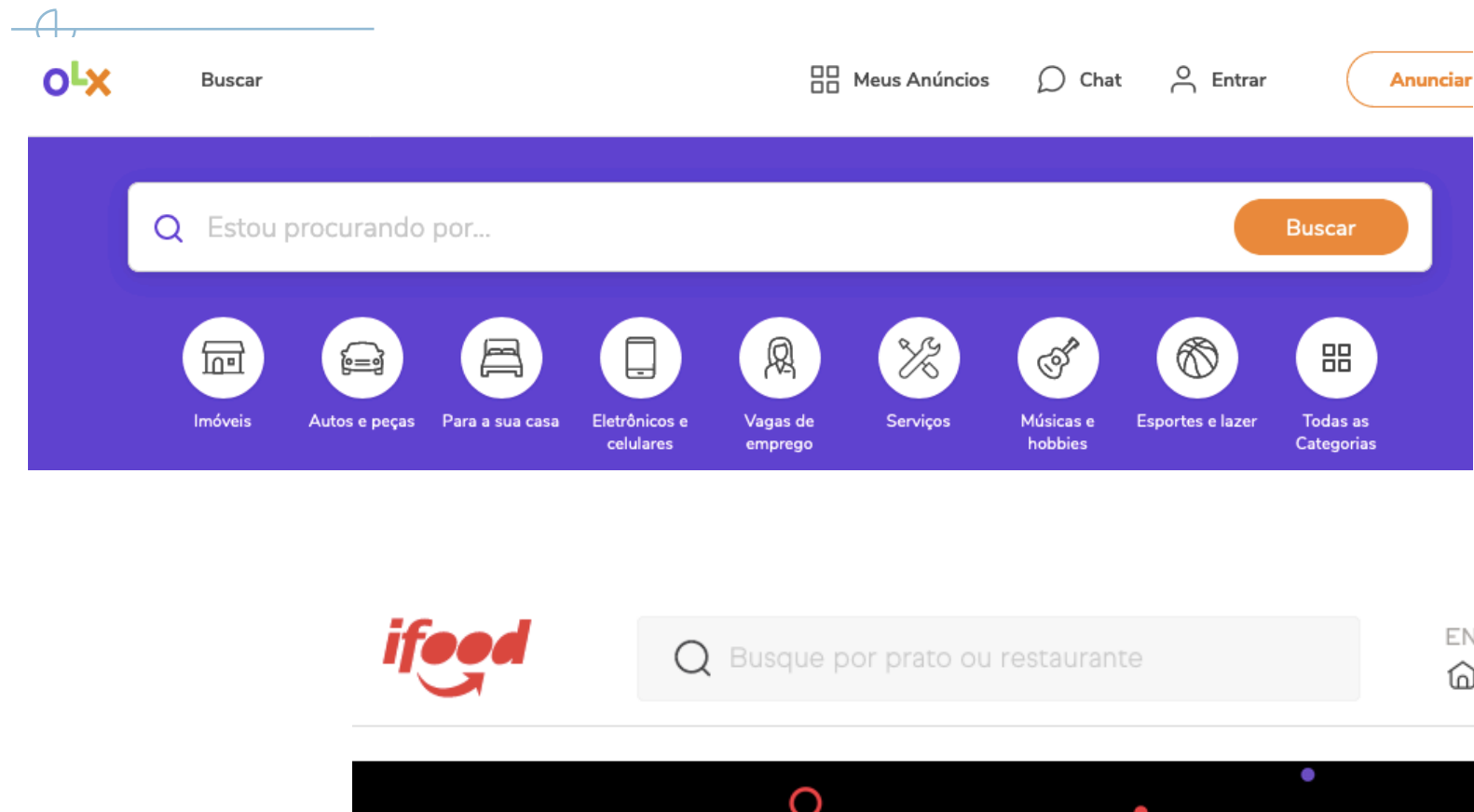
Netshoes (busca multitmodal)



Busca no Instagram



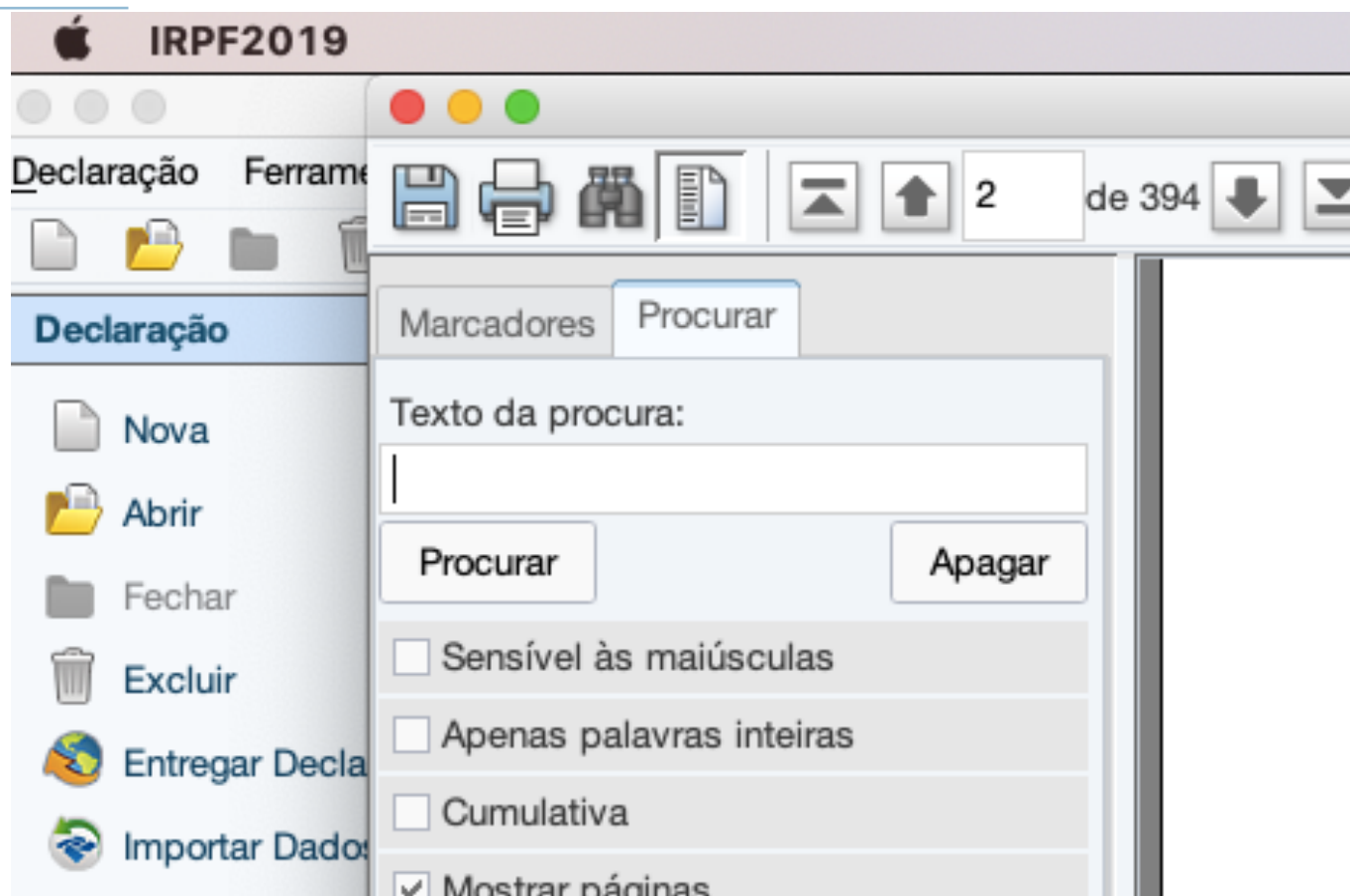
Busca no OLX, Ifood,...



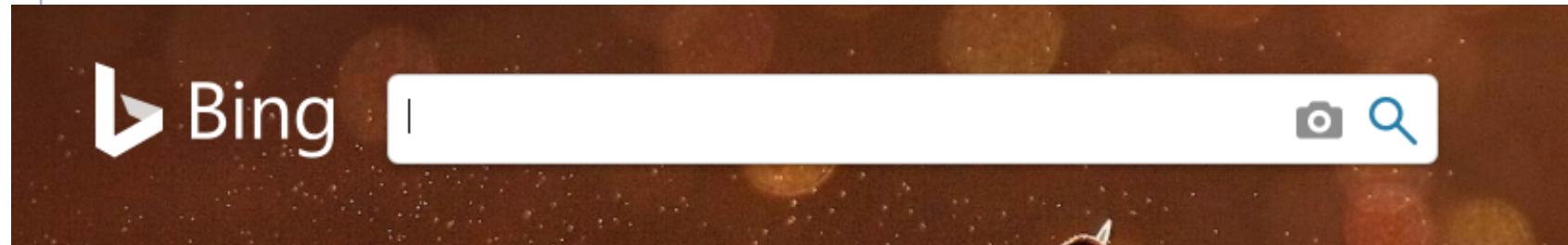
No Jusbrasil



Busca no Imposto de Renda



Busca em Máquinas de Busca



Google





Na wikipedia

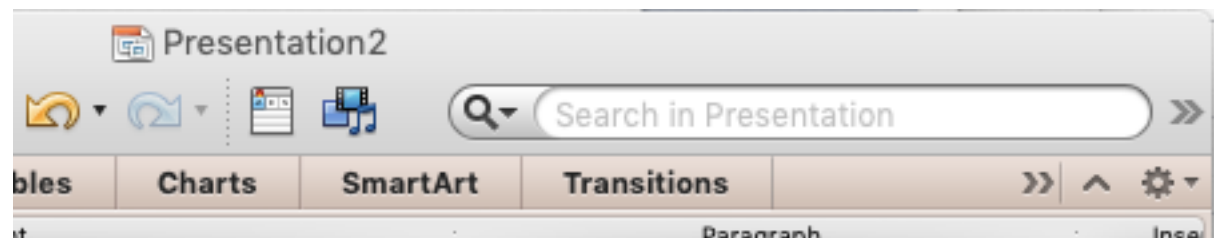
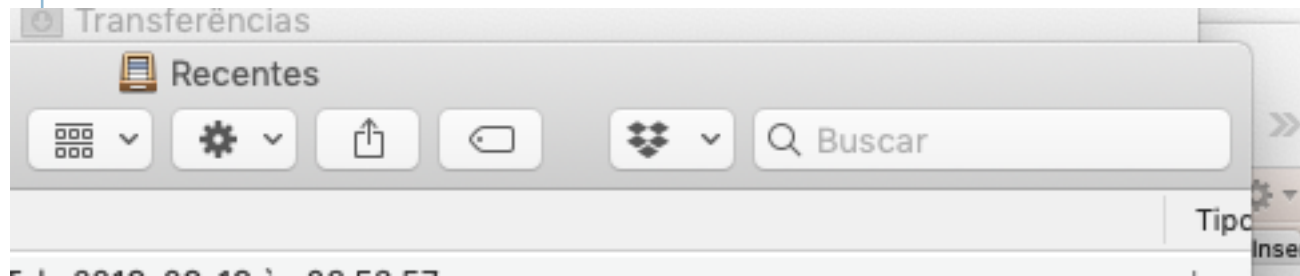


WIKIPEDIA
A enciclopédia livre

Português 1 013 000+ artigos	English 5 930 000+ articles
Español 1 545 000+ artículos	日本語 1 168 000+ 記事
Deutsch 2 343 000+ Artikel	Русский 1 568 000+ статей
Français 2 139 000+ articles	Italiano 1 552 000+ voci
中文 1 073 000+ 條目	Polski 1 359 000+ haseł

PT ▾

Até aqui no meu computador...



Pensando fora da caixa...

(de busca)

- Quando faço um sistema de recomendação, ele envolve muitos conceitos ligados a RI.
- Quando tenho que classificar documentos, dominar bem os conceitos de RI pode ajudar
- Quando tenho que fazer um crawler, conceitos de RI podem ajudar

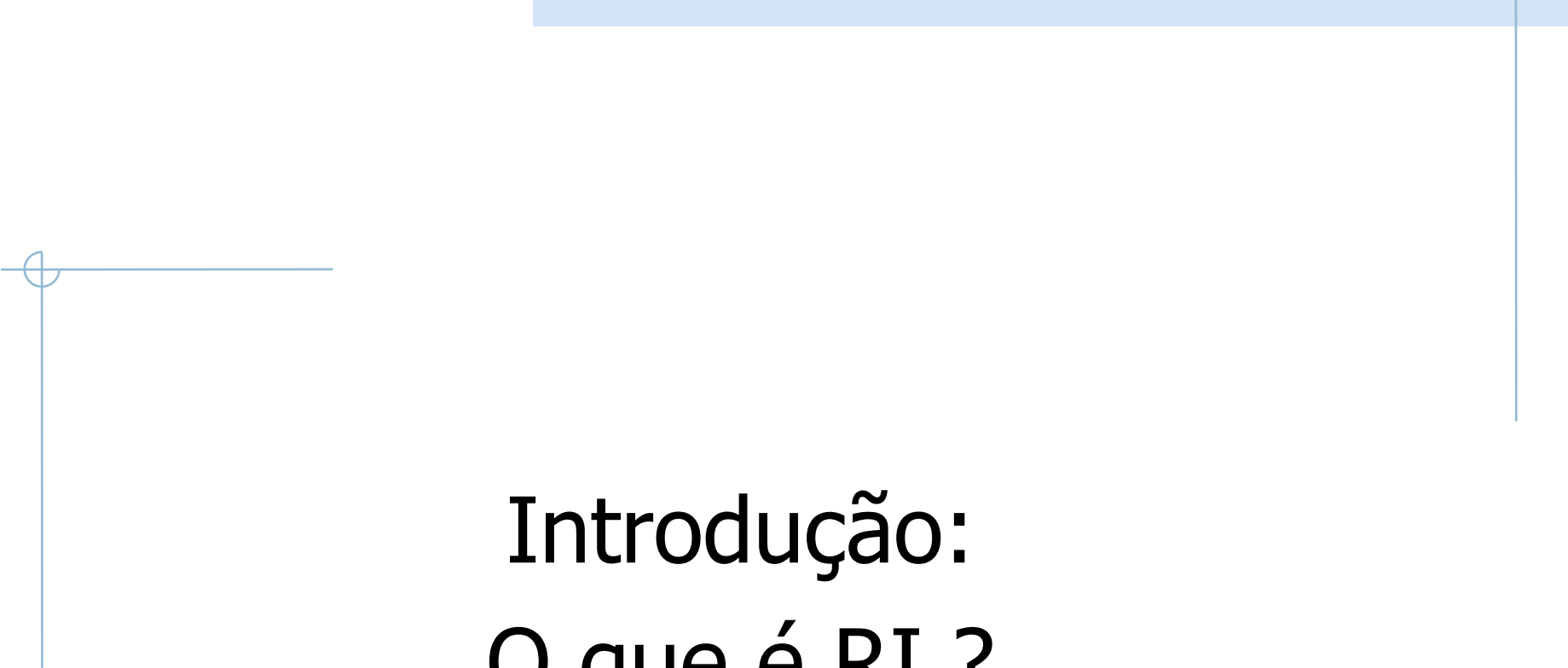
Pensando fora da caixa...

(de busca)

- Sistemas de segurança para sites
- Detecção de plágio
- Identificação de SPAM
- Identificação de autoria de textos

RESUMO:

- Conhecimento útil para muitos problemas práticos do dia-a-dia ligados à computação



Introdução:

O que é RI ?

O que é RI ?

- Recuperação de informação é uma subárea da ciência da computação que trata da recuperação automática de informação (normalmente contida em documentos textuais).

Documentos

- Fazem o papel de registros de dados.
- Podem conter qualquer tipo de mídia (texto, imagem, som).
- Normalmente são compostos de textos em linguagem natural ou de informação textual associada a outros tipos de dados.

Exemplos

(<https://rafaelluccasvsantana.jusbrasil.com.br/artigos/795241195/decreto-10189-19-indulto-natalino-violata-principio...>)

Princípio da Igualdade e da Isonomia

Indulto de Natal

Indulto



Decreto 10.189/19 (Indulto "Natalino") viola princípio da isonomia.

4 0 COMENTAR | 2

SALVAR f t e

Publicado por Rafael Santana

há 18 horas 58 visualizações



determinados crimes na modalidade culposa e pelo excesso culposos nos termos do artigo 23 do Código Penal e à militares das Forças Armadas, também pelo excesso culposos nos termos do artigo 45 do Código Penal Militar.

A Constituição Federal/1988 em seu artigo 84, XII, atribuiu como ato privativo do Presidente da República conceder o indulto, que nos termos do artigo 107, II, do Código Penal, é causa de extinção da punibilidade.

Conforme leciona Giuseppe Maggiore: “Uma das mais antigas formas de extinção da pretensão punitiva é a indulgência do príncipe, que se expressa em três instituições: a anistia, o indulto e a graça. A indulgentia principis se justifica como uma medida equitativa endereçada a suavizar a aspereza da justiça (supplementum iustitiae), quando particulares circunstâncias políticas, econômicas e sociais, fariam esse rigor aberrante e iníquo. Desse modo, atua como um ótimo meio de pacificação social”.

Outro exemplo:

<https://presrepublica.jusbrasil.com.br/legislacao/100760/decreto-3800-01...>



Edleno, esse documento é importante para os interessados em quais tópicos?

[Adicionar tópicos](#)

Decreto 3800/01 | Decreto nº 3.800, de 20 de abril de 2001

[Salvar](#) · [0 comentários](#) · [Imprimir](#) · [Reportar](#)

Publicado por [Presidência da Republica](#) (extraído pelo Jusbrasil) - 18 anos atrás

Ver artigo:

Regulamenta os arts. 40, 90 e 11 da Lei no 8.248, de 23 de outubro de 1991, e os arts. 80 e 11 da Lei no 10.176, de 11 de janeiro de 2001, que tratam do benefício fiscal concedida às empresas de desenvolvimento ou produção de bens e serviços de informática e automação, que investirem em atividades de pesquisa e desenvolvimento em tecnologia da informação, e dá outras providências.

(Revogado)

O VICE-PRESIDENTE DA REPÚBLICA

(Revogado)

O VICE-PRESIDENTE DA REPÚBLICA

(Revogado)

, no exercício do cargo de Presidente da República, usando da atribuição que lhe confere o art. 84, inciso IV, da

Constituição, e tendo em vista o disposto no art. 12 da Lei no 10.176, de 11 de janeiro de 2001, DECRETA:

(Revogado)

Art. 10 As empresas de desenvolvimento ou produção de bens e serviços de informática e automação, que investirem em atividades de pesquisa e desenvolvimento em tecnologia da informação, farão jus aos seguintes benefícios fiscais relativos ao Imposto sobre Produtos Industrializados - IPI incidentes sobre os bens de que trata o § 10 deste artigo, desde que atendidos os requisitos estabelecidos neste Decreto:

(Revogado)

I - nas regiões de influência da Superintendência do Desenvolvimento da Amazônia - SUDAM, da

Outro exemplo no site jusbrasil.com.br

2º Grau

Tribunal de Contas da União TCU - REPRESENTAÇÃO (REPR) : RP 02920620170

👍 0 👎 0

🔖 SALVAR    



Publicado por Tribunal de Contas da União

ano passado

RESUMO

INTEIRO TEOR

RELATÓRIO

VOTO

VOTO REVISOR

” EMENTA PARA CITAÇÃO

Processo	RP 02920620170
Órgão Julgador	Plenário
Partes	3. Representante: Ministério Público Federal/Procuradoria da República no Amazonas., Superintendência da Zona Franca de Manaus (Suframa).
Julgamento	12 de Dezembro de 2018
Relator	AUGUSTO SHERMAN

Ementa

REPRESENTAÇÃO. POSSÍVEIS IRREGULARIDADES NA APLICAÇÃO DE RECURSOS DE PROGRAMAS PRIORITÁRIOS DECORRENTES DE CONTRAPARTIDA DE EMPRESAS QUE USUFRUEM DE BENEFÍCIOS

Mais exemplos (doc no ifood)



Fast Temaki - P10

★ 4.5 • Japonesa • 2,2 km

28-38 min • Fechado



Enjoy Mais Que Temaki

★ - • Japonesa • 2,5 km

45-55 min • Fechado

Combo 20 hots e 2 temakis (10 hot philadélfia, 10 hot butterfly, 1 temaki philadelfia, 1 temaki butterfly)

10 hot philadélfia, 10 hot butterfly, 1 temaki philadelfia, 1 temaki butterfly

R\$ 60,00 ~~R\$ 80,00~~

 **Fest Sushi**

65-75 min

Documentos

- São as unidades de informação que buscamos:
- Anúncios num sites de anúncio
- Hotéis num site de hotéis
- Pratos e os restaurantes num site de alimentação
- Artefatos jurídicos num site para a área jurídica (lei, jurisprudência, artigos de advogados e etc...)

Recuperar Dados e Recuperar Informação

- Confusão entre recuperação de informação e recuperação de dados.
- Recuperação de Dados:
 - Tarefas precisas;
 - Sistemas não visam incorporar o significado do que está sendo buscado;
 - Respostas devem ser corretas;

Recuperar Dados e Recuperar Informação

- Recuperação de Informação:
 - Tarefas imprecisas;
 - Sistemas tentam modelar o significado do que está sendo buscado;
 - Objetivo é trazer as melhores respostas (normalmente não há o conceito de resposta correta);

Recuperar Dados e Recuperar Informação

- Todo sistema de recuperação de informação é também um sistema de recuperação de dados, mas o contrário nem sempre é verdade.

Exemplos

- Recuperação de Dados:
 - Uma busca por documentos que contém a palavra Manaus;
- Recuperação de Informação:
 - Uma busca por bons documentos que falam sobre a cidade de Manaus;



Dúvidas?



Histórico

Histórico

- Sistemas de RI utilizando computadores surgiram na década de 60.
- Principal objetivo era automatizar o acesso a informação em bibliotecas.

Principais focos até final dos anos 80

- RI em catálogos de bibliotecas.
- RI em Jornais, revistas e enciclopédias eletrônicas.
- RI em bases de dados de empresas.

Anos 90 e Anos 2000

- Recuperação de Informação na Web:
 - Abundância de informação não estruturada;
 - Publicação sem controle central e diversidade;
 - Dificuldade na busca de informação específica;

A Web

- Google tem mais de 10 hexabytes.
- Web deve ter alguns trilhões de páginas (CHUTE GROSSEIRO, número muito difícil de estimar hoje em dia)
- Mais chutes: Google usa mais de 1 milhão de máquinas para sua máquina de busca!!
- OUTRO CHUTE: Brasil deve ter mais de 1 centena de bilhões
- Tendência de crescimento continua

2010 em diante

- Mundo vive uma expansão a conectividade entre pessoas.
- Plataformas mobile, dispositivos que interagem entre si, com pessoas e com o ambiente
- Muitos problemas de busca/classificação onde precisamos avaliar as distâncias entre objetos.

Novos serviços online baseados em informação

- OLX
- GETNINJAS
- IFOOD
- JUSBASIL
- UBEREATS
- E-COMMERCE EM GERAL GANHA FORÇA E BUSCA É ESSENCIAL!

(muitas empresas no Brasil com oportunidades de emprego)

Amplo uso de machine learning (ML)

- Carros autônomos
- Deep Learning
- Atendentes virtuais e chatbots (Google, Apple e etc...)
- Máquinas substituindo motoristas, médicos e etc... ?
- Pode-se usar bastante ML em RI
- Conceitos de RI podem ajudar a melhorar muitas soluções de ML



Dúvidas?



Métricas para avaliação de Sistemas de RI

Avaliação de Sistemas de Busca

- N – conjunto de documentos relevantes identificados pelos especialistas
- R – conjunto de documentos respondidos pelo sistema que foram examinados.

Precisão e Revocação

$$\text{Precisão} = \frac{|N \cap R|}{|R|}$$

$$\text{Revocação} = \frac{|N \cap R|}{|N|}$$

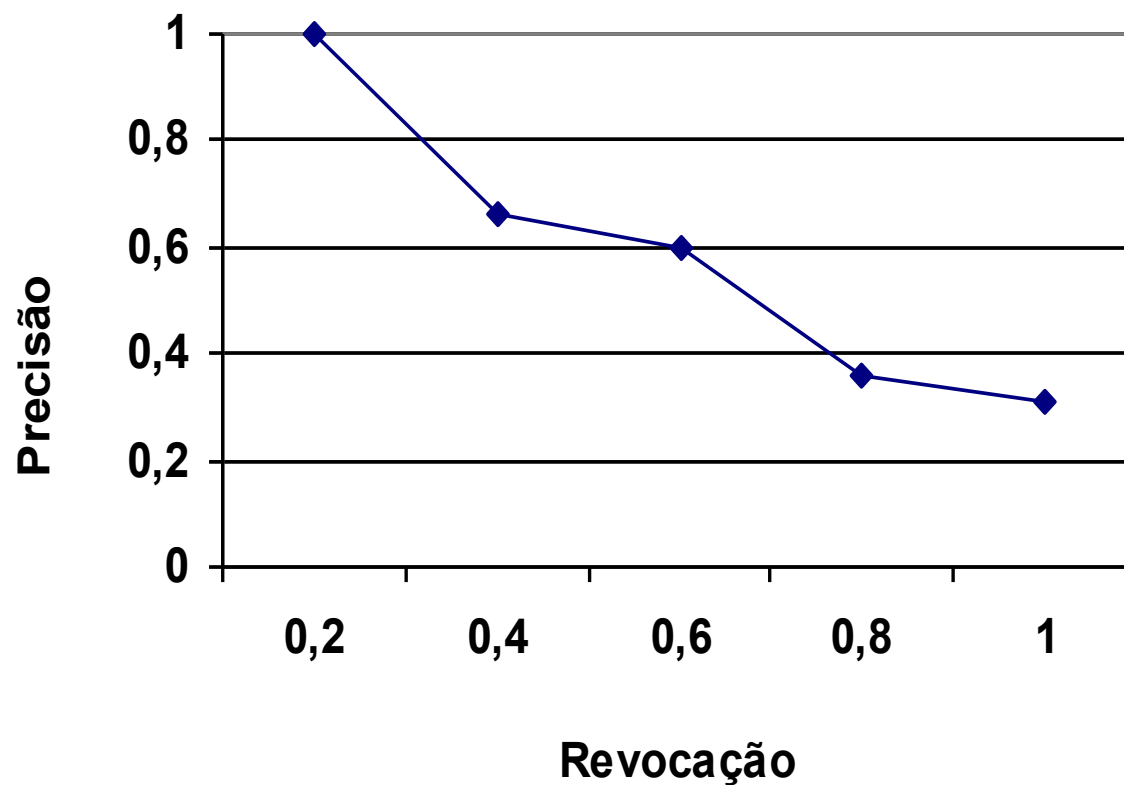
Curva de Precisão e Revocação

- Para facilitar a avaliação dos resultados é traçado um gráfico que mostra a evolução da precisão em função da revocação.
- Gráfico é conhecido como curva de precisão e revocação.

Exemplo

- Documentos relevantes: {1, 4, 8, 44, 72}.
- Um sistema recupera o vetor resultado: <8, 22, 72, 3, 1, 2, 24, 6, 33, 45, 4, 48, 55, 32, 11, 44>.
- O nível de revocação 20% é atingido quando encontramos o primeiro documento relevante (8), a precisão é de $1/1 = 100\%$.
- Para revocação de 40% a precisão é igual a $2/3 = 66\%$.

Curva de Precisão e Revocação

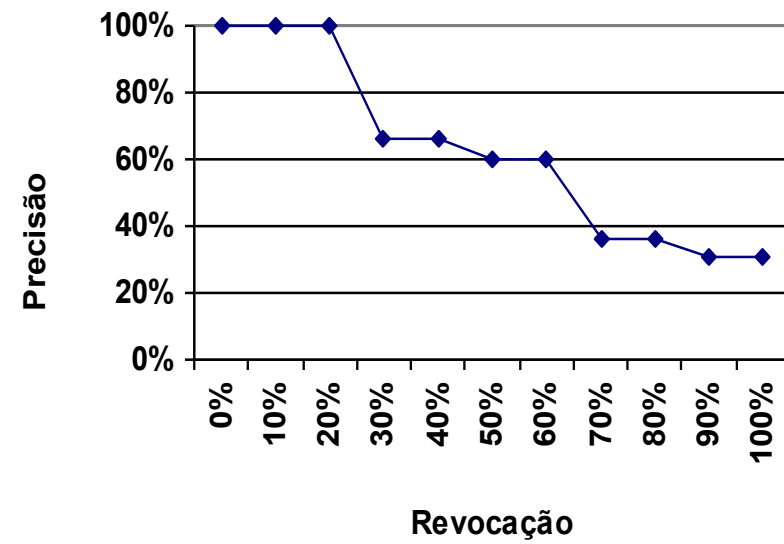
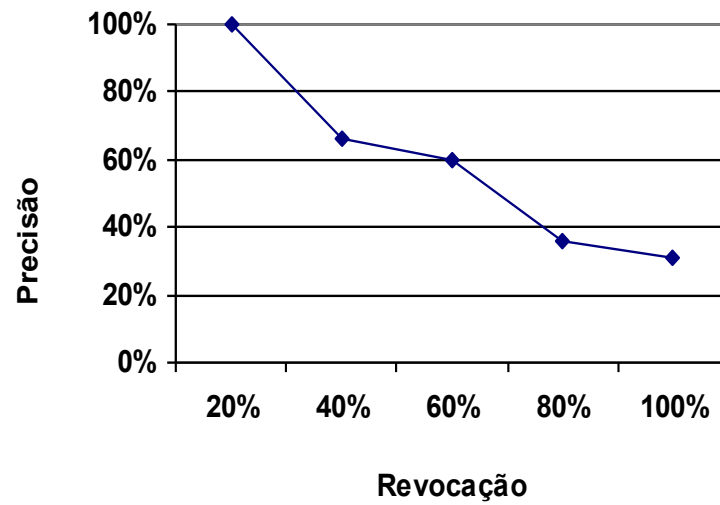


Precisão média

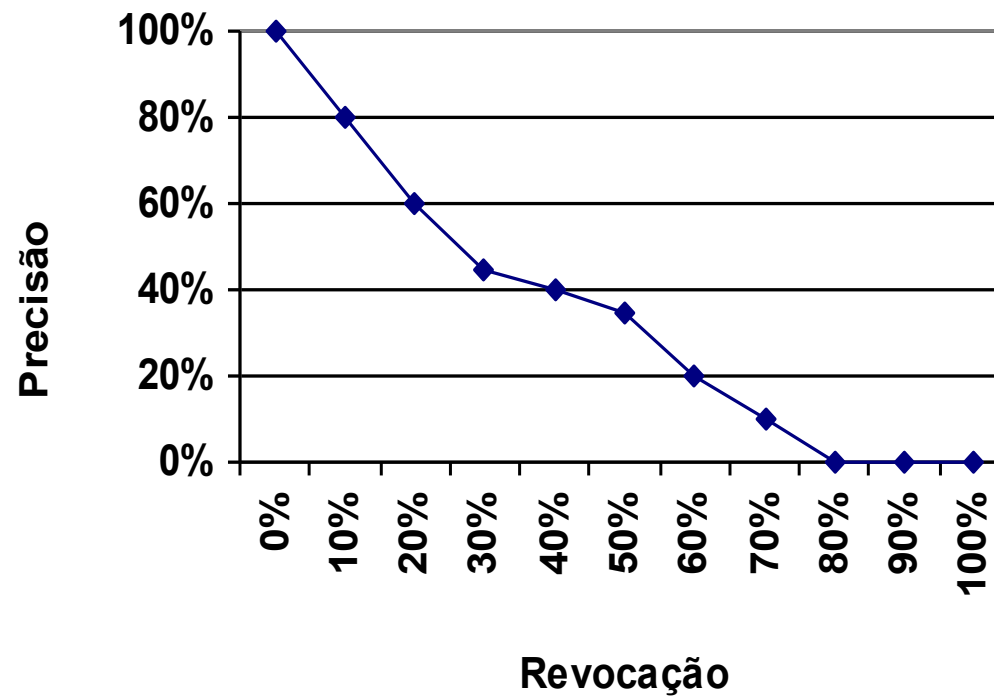
- Normalmente é interessante que se faça a avaliação do sistema utilizando-se uma média das precisões obtidas em várias consultas
- Pontos de precisão conhecidos são diferentes para cada consulta
 - Solução é criar uma forma de se ter valores conhecidos nos mesmos pontos em todas as consultas.

Precisão nos 11 pontos

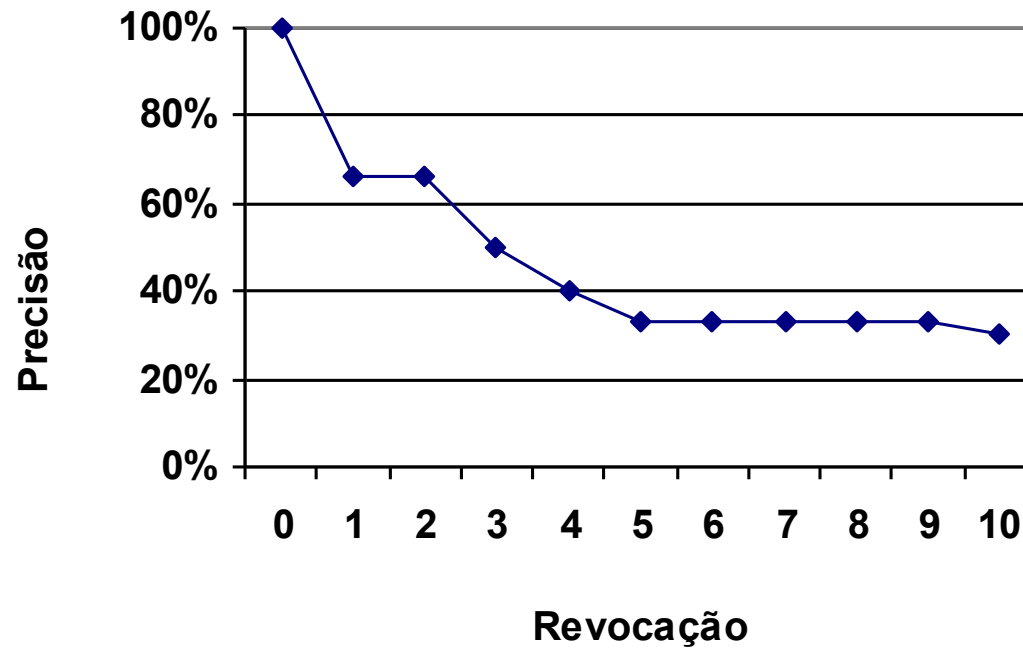
- Utiliza-se um método de interpolação para se obter a precisão em 11 pontos de revocação (0%, 10%, 20%, ..., 100%)
- Interpolação é feita tomando-se a precisão máxima conhecida entre o ponto atual e o próximo. Se não houver resultado, busca-se os próximos pontos até que se tenha uma definição



Se não recupera todos os relevantes precisão cai a zero



Precisão também pode ser computada em função do número de documentos vistos



Neste caso também utiliza-se interpolação

Exercício

- Documentos relevantes: {1, 4, 8, 25, 44, 53, 72}.
- Um sistema recupera o vetor resultado: <8, 22, 72, 1, 3, 2, 25, 6, 33, 45, 4, 48, 55, 32, 11, 44>.
- Mostre a curva de precisão e revocação utilizando os 11 pontos padronizados

Calculando valores únicos

- Em alguns casos é interessante que se tenha um único valor de precisão para cada consulta
- Com este valor é possível comparar diretamente dois sistemas e determinar qual o melhor

Medidas de precisão escalares

- Precisão média nos relevantes encontrados
 - Tira-se a média das precisões nos pontos de revocação onde apareceram documentos relevantes (MAP não interpolado)
 - Há também o **MAP Interpolado** (média nos 11 pontos) (**CAIU EM DESUSO!**)

Exercício

- Calcule o MAP para o exercício anterior

Medidas de precisão escalares

- Precisão@R
 - Calcula-se a precisão na R-ésima posição do ranking
 - Muito utilizada quando assume-se que usuário está interessado nos R primeiros.
 - Por exemplo, nas máquinas costuma-se assumir que usuário está interessado em respostas apenas entre os 10 primeiros itens

- Réo nú

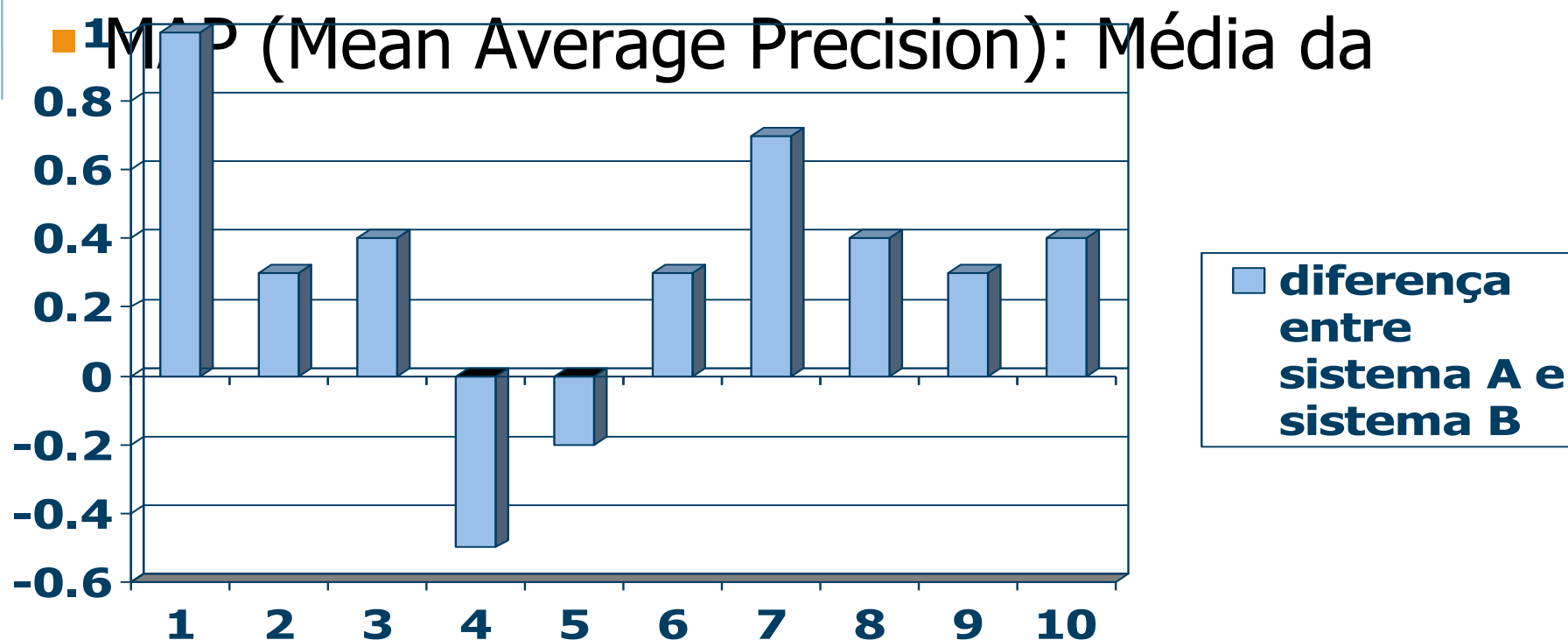
- R é o número de relevantes
- Irrelevant_R(r) é o número de documentos irrelevantes acima de r, entre os R+10 documentos do topo (valor maximo é R+10).
- Falha quando número de relevantes é pequeno
- Ex :
$$\frac{1}{7} \times ((1 - 0/17) + (1 - 1/17) + (1 - 1/17) + (1 - 1/17) + (1 - 3/17) + (1 - 6/17) + (1 - 10/17) + 0) = 0.67$$

Exercício

Cacule o Bpref-10 para o exercício anteriorl

Histograma de precisão

- Pode-se montar um histograma com as diferenças entre dois sistemas em vários pontos de precisão-R



Resumos comparativos

- Pode-se ainda montar resumos comparativos sobre os sistemas que estão sendo experimentados
 - Exemplo: número de consultas usadas no experimento, número médio de docs recuperados por consulta, número médio de relevantes por consulta, precisão-10 de cada sistema e assim por diante.

Medidas alternativas

- Medida F1(Média Harmônica)

$$F1(j) = \frac{2}{\frac{1}{R(j)} + \frac{1}{P(j)}}$$

Onde $R(j)$ é a revocação em um dado ponto j e $P(j)$ é a precisão neste ponto

A medida F é útil para combinar a precisão e a revocação em um único número

Medidas Alternativas

- Medida F

$$F(j) = \frac{1 + b^2}{\frac{b^2}{R(j)} + \frac{1}{P(j)}}$$

- Valores de b maiores que 1 indicam que o usuário está mais interessado na precisão
- Valores menores que 1 indicam que usuário está mais interessado na revocação

Exercício

- Calcule a medida F para o exercício anterior

MRR(Mean Reciprocal Ranking)

$$MRR(S, Q) = \frac{\sum_{q \in Q} \frac{1}{PosRel(S(q))}}{|Q|}$$

- Onde Q é um conjunto de consultas
- S é um sistema de ranking
- $PosRel(R(q))$ é a posição da primeira resposta relevante no ranking do sistema S para a consulta q
- $|Q|$ é o número de consultas avaliadas

Exercício

- Calcule o MRR para o exercício anterior

Métricas baseadas em Ganho Cumulativo (CG)

- Quando examinamos uma resposta de um sistema, fica claro que:
 - Alguns documentos relevantes atendem melhor às necessidades dos usuários que outros
 - Quanto mais longe do topo um documento relevante está, menor a sua utilidade na resposta
- Métricas baseadas em ganho cumulativo tentam incorporar estes dois fatos na avaliação

Ganho Cumulativo (CG)

- Documentos da resposta são substituídos pelos seus graus de relevância:

$\langle 3, 0, 1, 2, 3, 0, 0, 1 \rangle$

- Ganho cumulativo (CG) é igual a soma dos valores de ganho obtidos até cada posição:

$\langle 3, 3, 4, 6, 9, 9, 9, 10 \rangle$

- CG leva em consideração a relevância, mas não a posição

Calcule do CG para o exemplo abaixo ate o 6o documento

- Documentos relevantes: $\{1(3), 4(1), 8(2), 25(1), 44(3), 53(3), 72(2)\}$.
- Um sistema recupera o vetor resultado: $\langle 8, 22, 72, 1, 3, 2, 25, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.
- Um sistema recupera o vetor resultado: $\langle 25, 22, 4, 3, 72, 2, 8, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.

Ganho Cumulativo Descontado (DCG)

- Inclui a noção de que documentos relevantes têm a utilidade reduzida na medida em que são apresentados mais longe do topo da resposta
- A proposta é incluir um fator de desconto no ganho de acordo com a posição na qual os documentos são apresentados
- Uma proposta é dividir pelo log da posição no ranking
- A base do logaritmo ajusta o fator de desconto e o log não é aplicado para a primeira posição do ranking

DCG

- Usando log na base 2:
 $\langle 3 ; 3 ; 2,51 ; 3 \dots \rangle$
- Valores médios de DCG podem ser computados para avaliar o desempenho de um sistema e gráficos de DCG podem ser criados para facilitar a visualização dos resultados da avaliação

Calcule do DCG para o exemplo abaixo ate o 6o documento

- Documentos relevantes: $\{1(3), 4(1), 8(2), 25(1), 44(3), 53(3), 72(2)\}$.
- Um sistema recupera o vetor resultado: $\langle 8, 22, 72, 1, 3, 2, 25, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.
- Um sistema recupera o vetor resultado: $\langle 25, 22, 4, 3, 72, 2, 8, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.

NDCG (Jarvelin et al, TOIS, 2002)

- Normalized Cumulative Discount Gain
- Ganho Cumulativo Descontado Normalizado
- Valor ótimo de DCG poderia ser obtido com sistema que coloca os documentos ordenados de forma decrescente por valor de relevância:
<3,3,3,3,2,2,2,2,1,1,1,0,0,0,0>
- NDCG divide o DCG de cada sistema pelo DCG de um sistema ideal, obtendo valores entre 0 e 1 para cada posição do ranking.
- NDCG@K: NDCG obtido na k-ésima posição do ranking.

Calcule do NDCG para o exemplo abaixo ate o 6o documento

- Documentos relevantes: $\{1(3), 4(1), 8(2), 25(1), 44(3), 53(3), 72(2)\}$.
- Um sistema recupera o vetor resultado: $\langle 8, 22, 72, 1, 3, 2, 25, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.
- Um sistema recupera o vetor resultado: $\langle 25, 22, 4, 3, 72, 2, 8, 6, 33, 45, 4, 48, 55, 32, 11, 44 \rangle$.

E depois de tirar a média

- 1 – Pode-se usar testes estatísticos para verificar se a diferença entre as médias é estatisticamente significativa
- 2 – Pode-se calcular a variância em torno da média com algum intervalo de confiança: tipicamente 95% ou 99%, e verificar se as duas médias estão distantes o suficiente

Testes estatísticos

- Hipótese nula: Hipótese de que dois resultados comparados são iguais
- Hipótese alternativa: Há diferenças entre os dois resultados (bicaudal)
- Testes estatísticos servem para dizer se a hipótese nula está descartada com um certo grau de certeza (p-value), normalmente entre 95% e 99%

Testes estatísticos

- Se o teste de significância falha, isso não quer dizer que os dois sistemas produzem resultados iguais. Isso quer dizer que seu experimento não foi conclusivo!
- Pode-se ampliar número de amostras para tornar experimento mais conclusivo
- Diferença ser considerada estatisticamente significativa não significa que diferença é necessariamente grande!!!!

Testes estatísticos

- Tentam estimar se os resultados de comparação entre dois sistemas foi obtido ao acaso (por sorte) ou se há uma diferença consistente
- Podem ser aplicados sobre as diversas métricas, tais como MAP, bpref, NDCG@K...

Há críticas sobre estes testes...

- Alguns autores da área de estatística criticam o uso deste tipo de testes e sugerem o uso de intervalos de confiança ao invés de testes estatísticos

T-test

- Proposto em 1908
- Autor (William S. Gosset) usou um codinome (student) porque não podia ser identificado. Por isso, o teste é conhecido como “t-student test”
- Uso em RI:
 - Utilizado para verificar se a hipótese “a média de valores de duas populações é igual” é válida ou não.

T-teste é paramétrico

- REQUISITOS:
- As duas amostras devem ter distribuição normal. Aplicar um teste de normalidade
- As duas amostras devem ter mesma variância. Aplicar um teste F de igualdade de variâncias.
- As duas amostras devem ser independentes.

Tipos de ttest

- Pareado e não pareado: quando queremos usar em RI são **pareados(ou emparelhados)**, as amostras representam dois momentos distintos para as mesmas consultas (mesma população)
- Não pareados são testes em que as amostras são diferentes. Não costumamos aplicar em RI

Tipos de ttest

- Bicaudal ou unicaudal: Em RI queremos **bicaudal** (two-sided ou two-tails). Responde se há ou não diferenças entre duas médias.
- Unicaudal: responde se diferença é positiva ou negativa em relação a uma média. (não usado em RI)

T-test

- Média das diferenças sobre desvio padrão multiplicado pela raiz do número de amostras.

$$t = \frac{\sum |a_i - b_i|}{\sqrt{\frac{N(\sum (a_i - b_i)^2) - (\sum |a_i - b_i|)^2}{N - 1}}}$$

- Resultado obtido é comparado ao valor da tabela de teste t, com coluna dada pelo grau de certeza esperado e linha dada pelo número de elementos comparados menos o número de amostras (N -1)

ϵ	.400	.250	.100	.050	.025	.010	.005	.001
n								
1	.325	1,000	3.078	6.314	12.706	31.821	63.656	318.289
2	.289	.816	1.886	2.920	4.303	6.965	9.925	22.328
3	.277	.765	1.638	2.353	3.182	4.541	5.841	10.214
4	.271	.741	1.533	2.132	2.776	3.747	4.604	7.173
5	.267	.727	1.476	2.015	2.571	3.365	4.032	5.894
6	.265	.718	1.440	1.943	2.447	3.143	3.707	5.208
7	.263	.711	1.415	1.895	2.365	2.998	3.499	4.785
8	.262	.706	1.397	1.860	2.306	2.896	3.355	4.501
9	.261	.703	1.383	1.833	2.262	2.821	3.250	4.297
10	.260	.700	1.372	1.812	2.228	2.764	3.169	4.144
11	.260	.697	1.362	1.796	2.201	2.718	3.106	4.075

- Calcule o resultado do teste t para:

S1	S2	S1-S2	s1-s2 ^2
35	29	6	36
50	40	10	100
10	10	0	0
40	45	5	25
21	20	1	1
19	15	4	16
70	60	10	100
30	25	5	25
80	70	10	100

$$t = \frac{\sum |a_i - b_i|}{\sqrt{\frac{N(\sum (a_i - b_i)^2) - (\sum |a_i - b_i|)^2}{N - 1}}}$$

Wilcoxon

- Proposto por Frank Wilcoxon, 1945
- Não paramétrico (uso mais amplo que t-test mas menos conclusivo)
- Mesmas questões (pareado, não pareado)



Muito além das métricas...

Coleções de referência

- Métricas são importantes, mas antes de calculá-las, precisamos ter conjuntos de consultas e avaliações de relevância para tais conjuntos

Coleções de referência

- A forma como se cria coleções tem sido objeto de estudo desde o início da área
- IR lida com incerteza, é difícil julgar uma resposta como boa. Muitas vezes é difícil saber qual a melhor resposta entre dois sistemas

Algumas premissas básicas

- Ao criar uma coleção de referência:
 1. Queremos medir a habilidade do sistema para recuperar os documentos relevantes para o usuário
 2. Habilidade de separar relevantes de não relevantes
 3. Facilidade para o usuário especificar o que deseja
 4. Descobrir se e quanto um sistema X é melhor ou não que um sistema Y.

Passos importantes

- Seleção das consultas:
 - De onde serão tiradas as consultas ? Elas representam uma boa amostra do que é/ será buscado no sistema ?
 - Há algum tipo de classificação importante entre as consultas: difíceis, fáceis, atemporais, temporais,
 - Quantas consultas devemos ter ?

Passos importantes

- Quem vai avaliar as respostas ?
 - Grau de imparcialidade dos avaliadores.
 - Representam os usuários que terão acesso ao sistema ? (entendem suas necessidades e como eles pensam ?)
 - São imparciais na sua avaliação ?
 - Entendem dos objetivos do sistema proposto ?
 - Quantos vão avaliar cada resposta ?

Passos importantes

- Que respostas vou mostrar para os avaliadores ?
 - Costuma-se mostrar a união do topo de um conjunto de sistemas. Depois toma-se a avaliação das respostas para obter-se o conjunto de relevantes
 - Problema: Se um novo sistema traz novas respostas relevantes muito diferentes das avaliadas, o mesmo vai ser prejudicado nas avaliações!

Passos importantes

- Que respostas vou mostrar para os avaliadores ?
 - Se mostro um conjunto muito pequeno, avaliação pode ficar incopleta
 - Se mostro um conjunto muito grande, avaliação pode ficar muito cara

TREC

- Text Retrieval Conference: conferência anual que cria e atualiza diversas coleções de referência para disponibiliza-las para pesquisadores avaliarem seus sistemas
- Como funciona: você pode inscrever seu grupo de pesquisa em uma task (tarefa) e submeter sistemas que farão parte da seleção de resultados avaliados
- Durante a tarefa, cada grupo recebe atividades que deve obrigatoriamente executar

TREC

- Ao final do ano, são publicados trabalhos mostrando os resultados obtidos pelos melhores sistemas dentro os que participaram da task.
- Mais que isso, as avaliações de consultas são disponibilizadas na forma de coleções de referência

TREC

- Ao final do ano, são publicados trabalhos mostrando os resultados obtidos pelos melhores sistemas dentro os que participaram da task.
- Mais que isso, as avaliações de consultas são disponibilizadas na forma de coleções de referência

TREC 2020

- <https://trec.nist.gov/pubs/call2020.html>

View on GitHub 

TREC-2019-Deep-Learning

Website for the TREC Deep Learning Track 2019

TREC 2019 Deep Learning Track Guidelines

Timetable

- August 7: Submissions close for document ranking task
- August 14: Submissions close for passage ranking task
- August 21: Optional Docker images due for the Replicable Runs Initiative
- November 13-15: TREC conference

Please see [recent announcement about updated deadlines for submitting runs for the passage](#)

Para discutir...

- Como comparar Google x Bing ??
- Problemas com métricas de avaliação
 - Subjetividade e contexto
 - Uso de logs x uso de consultas específicas
 - Níveis de relevância