Universidade Federal do Rio de Janeiro (UFRJ)
Departamento de Ciência da Computação (DCC)

DEPARTAMENTO DE
CIÊNCIA DA
COMPUTAÇÃO
- UFRJ -

Recuperação da Informação (MAB605)

Avaliação da Recuperação

Profa. Giseli Rabello Lopes

Roteiro

- Introdução
- Precisão e Revocação
- Sumários com um único valor
- *nDCG*
- Correlação de *rankings*
- Referências

Introdução [Baeza-Yates & Ribeiro Neto, 2013]

- **Avaliação da recuperação** é um componente crítico e integrante de qualquer sistema moderno de RI
  - Avaliar um sistema de RI é medir quão bem um sistema atende às necessidades de informação dos usuários
  - Sem uma avaliação adequada da recuperação, não se pode
    - Determinar o quão bem um sistema de RI está executando
    - Comparar a *qualidade* do sistema de RI com a de outros sistemas, objetivamente

Introdução [Baeza-Yates & Ribeiro Neto, 2013]

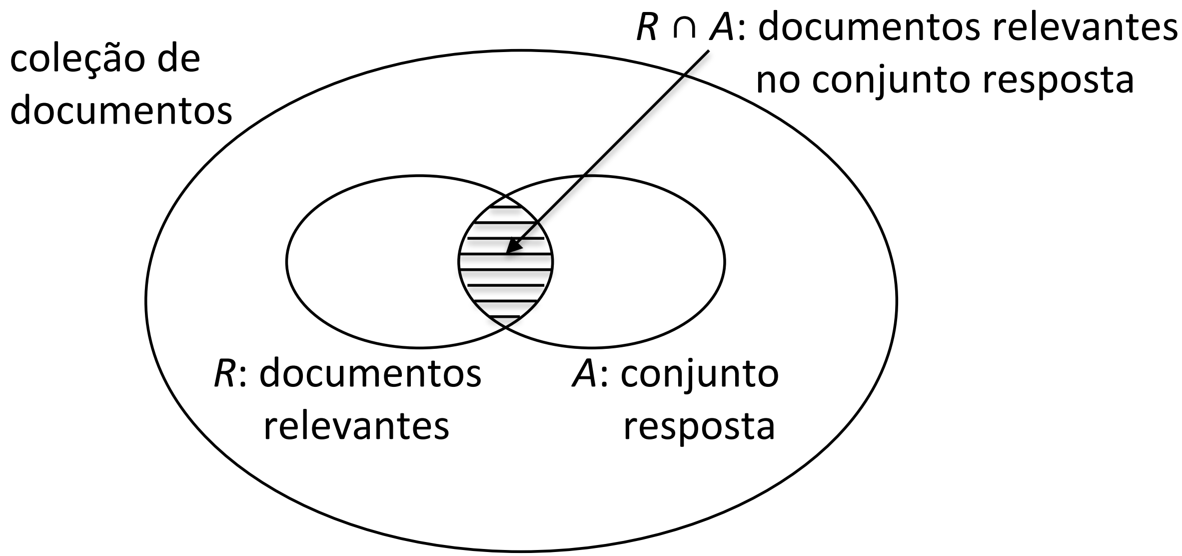
- A avaliação sistemática de um sistema de RI permite responder questões tais como:
  - Uma modificação da função de *ranking* é proposta, devemos ir adiante e implementá-la?
  - Uma nova função probabilística de ranqueamento foi projetada, ela é superior aos *rankings* do modelo vetorial e BM25?
  - Para quais tipos de consultas uma dada modificação no *ranking* funciona melhor?
- A falta de avaliação impede responder essas questões e impossibilita o ajuste da função de ranqueamento

Introdução [Baeza-Yates & Ribeiro Neto, 2013]

- *Avaliação da recuperação* consiste em associar uma métrica quantitativa aos resultados produzidos por um sistema de RI
- *Avaliação da recuperação* é avaliar a *qualidade* dos resultados, *não o desempenho* do sistema em termos do tempo de processamento das consultas

Precisão e Revocação
[Baeza-Yates & Ribeiro-Neto, 2013]

- Considere,
  - I: uma requisição de informação
  - R: o conjunto de documentos relevantes para I
  - A: o conjunto resposta para I, gerado por um sistema de RI
  - R ∩ A: a intersecção entre os conjuntos R e A



Precisão e Revocação
[Baeza-Yates & Ribeiro-Neto, 2013]

- **Revocação** é a fração de documentos relevantes (conjunto *R*) que foram recuperados

$$\boxed{Recall = |R \cap A| / |R|}$$

- **Precisão** é a fração de documentos recuperados (conjunto *A*) que são relevantes

$$\boxed{Precision = |R \cap A| / |A|}$$

Precisão e Revocação
[Baeza-Yates & Ribeiro-Neto, 2013]

- O usuário não é normalmente apresentado a todos os documentos no conjunto de resposta *A* de uma só vez
  - Usuário vê um conjunto de documentos ranqueados e examina-os a partir do topo
- Então, precisão e revocação variam conforme o usuário prossegue com sua análise do conjunto *A*
- O mais apropriado é, então, traçar uma curva de **precisão *versus* revocação**

Exemplo – Precisão e Revocação

- Considere uma coleção de referência e um conjunto de consultas de teste
- Sejam R_{q1} e R_{q2} os conjuntos de documentos relevantes para as consultas q_1 e q_2, respectivamente:
  - $R_{q1} = \{d_3, d_5, d_9\}$ e $R_{q2} = \{d_1, d_2, d_6, d_9\}$
- Considere os *rankings* gerados por um algoritmo de RI para as consultas q_1 e q_2 (documentos relevantes estão marcados com **v**) a seguir

Exemplo – Precisão e Revocação

Ranking para q_1

#	A
1	d_1
2	d_3 v
3	d_5 v
4	d_9 v
5	d_2

$R_{q1} = \{d_3, d_5, d_9\}$

Ranking para q_2

#	A
1	d_9 v
2	d_3
3	d_4
4	d_1 v
5	d_2 v

$R_{q2} = \{d_1, d_2, d_6, d_9\}$

Exemplo – Precisão e Revocação

Ranking para q_1

#	A
1	d_1
2	d_3 v
3	d_5 v
4	d_9 v
5	d_2

$$R_{q1} = \{d_3, d_5, d_9\}$$

Recall	Precision
0,33	0,5
0,67	0,67
1,0	0,75

Ranking para q_2

#	A
1	d_9 v
2	d_3
3	d_4
4	d_1 v
5	d_2 v

$$R_{q2} = \{d_1, d_2, d_6, d_9\}$$

Recall	Precision
0,25	1,0
0,5	0,5
0,75	0,6

Curva de precisão interpolada para os 11 níveis padrão de revocação

- Seja $r_j, j \in \{0,1,2,\ldots,10\}$, uma referência ao *j*-ésimo nível padrão de revocação
- Então, $P(r_j) = \max_{\forall r \mid rj \leq r} P(r)$

> Precisão interpolada para um nível de revocação *j* será o valor máximo de precisão encontrado para qualquer nível de revocação maior ou igual *j*.

Exemplo – Precisão e Revocação

Ranking para q_1

#	A
1	d_1
2	d_3 v
3	d_5 v
4	d_9 v
5	d_2

$$R_{q1} = \{d_3, d_5, d_9\}$$

Recall	Precision
0,33	0,5
0,67	0,67
1,0	0,75

Ranking para q_2

#	A
1	d_9 v
2	d_3
3	d_4
4	d_1 v
5	d_2 v

$$R_{q2} = \{d_1, d_2, d_6, d_9\}$$

Recall	Precision
0,25	1,0
0,5	0,5
0,75	0,6

Exemplo – Precisão e Revocação

Ranking para q_1

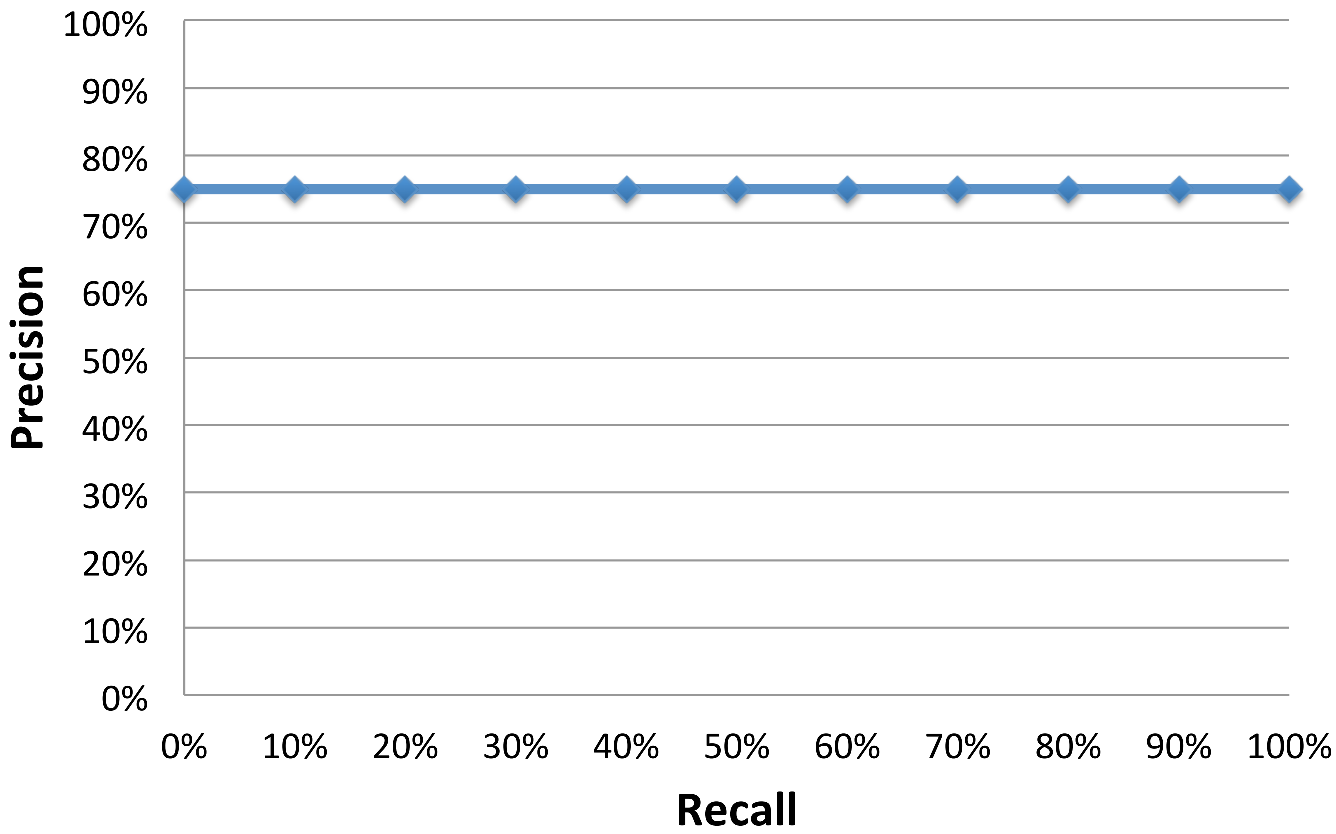
#	A
1	d_1
2	d_3 v
3	d_5 v
4	d_9 v
5	d_2

$R_{q1} = \{d_3, d_5, d_9\}$

Recall	Precision
0,33	0,5
0,67	0,67
1,0	0,75

Para q_1

Recall	Precision
0,0	0,75
0,1	0,75
0,2	0,75
0,3	0,75
0,4	0,75
0,5	0,75
0,6	0,75
0,7	0,75
0,8	0,75
0,9	0,75
1,0	0,75



Recall	Precision
0,33	0,5
0,67	0,67
1,0	0,75

Precisão interpolada para um nível de revocação *j* será o valor máximo de precisão encontrado para qualquer nível de revocação maior ou igual *j*.

Exemplo – Precisão e Revocação

Recall	Precision
0,25	1,0
0,5	0,5
0,75	0,6

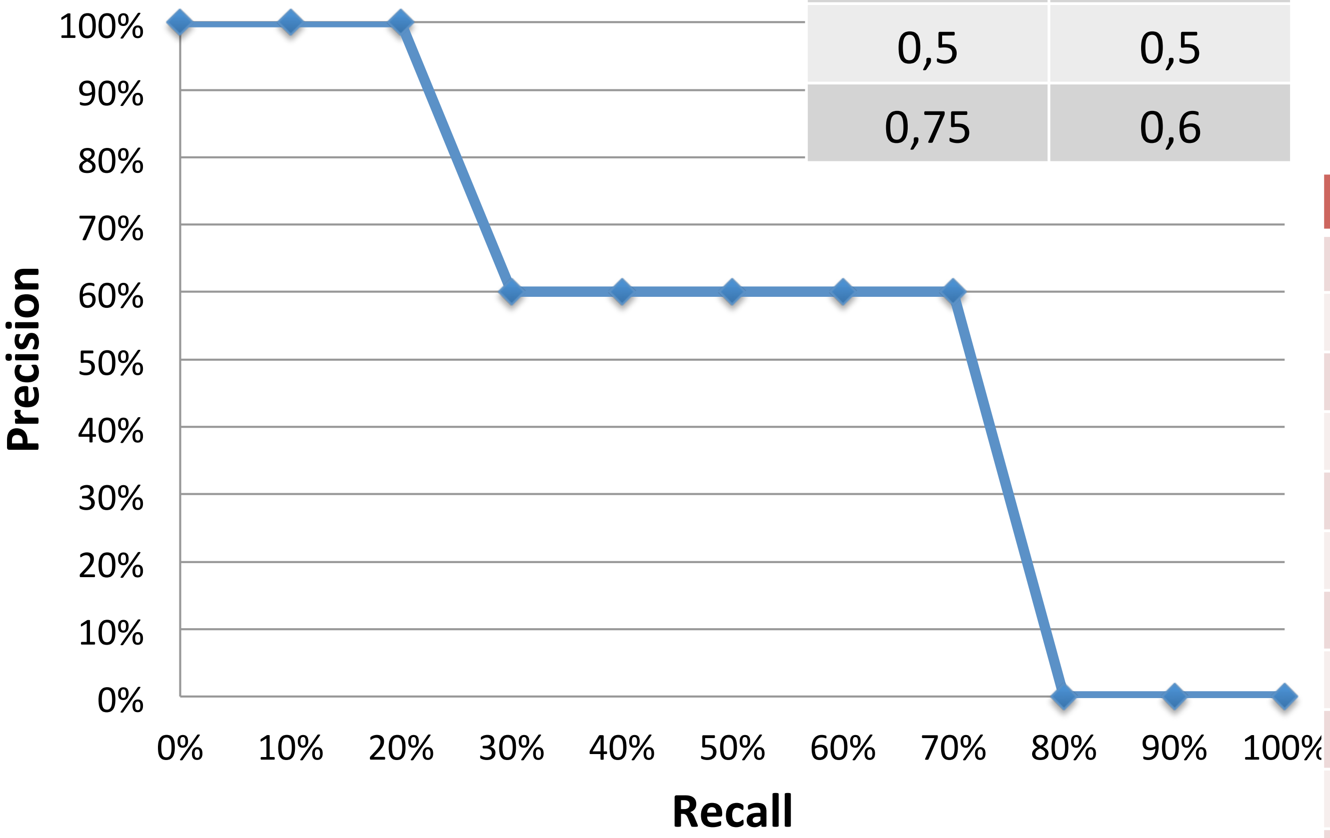
Ranking para q_2

#	A
1	d_9 ν
2	d_3
3	d_4
4	d_1 ν
5	d_2 ν

$$R_{q2} = \{d_1, d_2, d_6, d_9\}$$

Precisão interpolada para um nível de revocação *j* será o valor máximo de precisão encontrado para qualquer nível de revocação maior ou igual *j*.

Recall	Precision
0,25	1,0
0,5	0,5
0,75	0,6



Para q_2

Recall	Precision
0,0	1,0
0,1	1,0
0,2	1,0
0,3	0,6
0,4	0,6
0,5	0,6
0,6	0,6
0,7	0,6
0,8	**0,0**
0,9	**0,0**
1,0	**0,0**

Precisão e Revocação

[Baeza-Yates & Ribeiro-Neto, 2013]

- Geralmente, os algoritmos de recuperação são avaliados sobre diversas consultas de teste
- Para avaliar a *qualidade* de recuperação para N_q consultas, calculamos a média das precisões para cada nível de revocação, como segue:

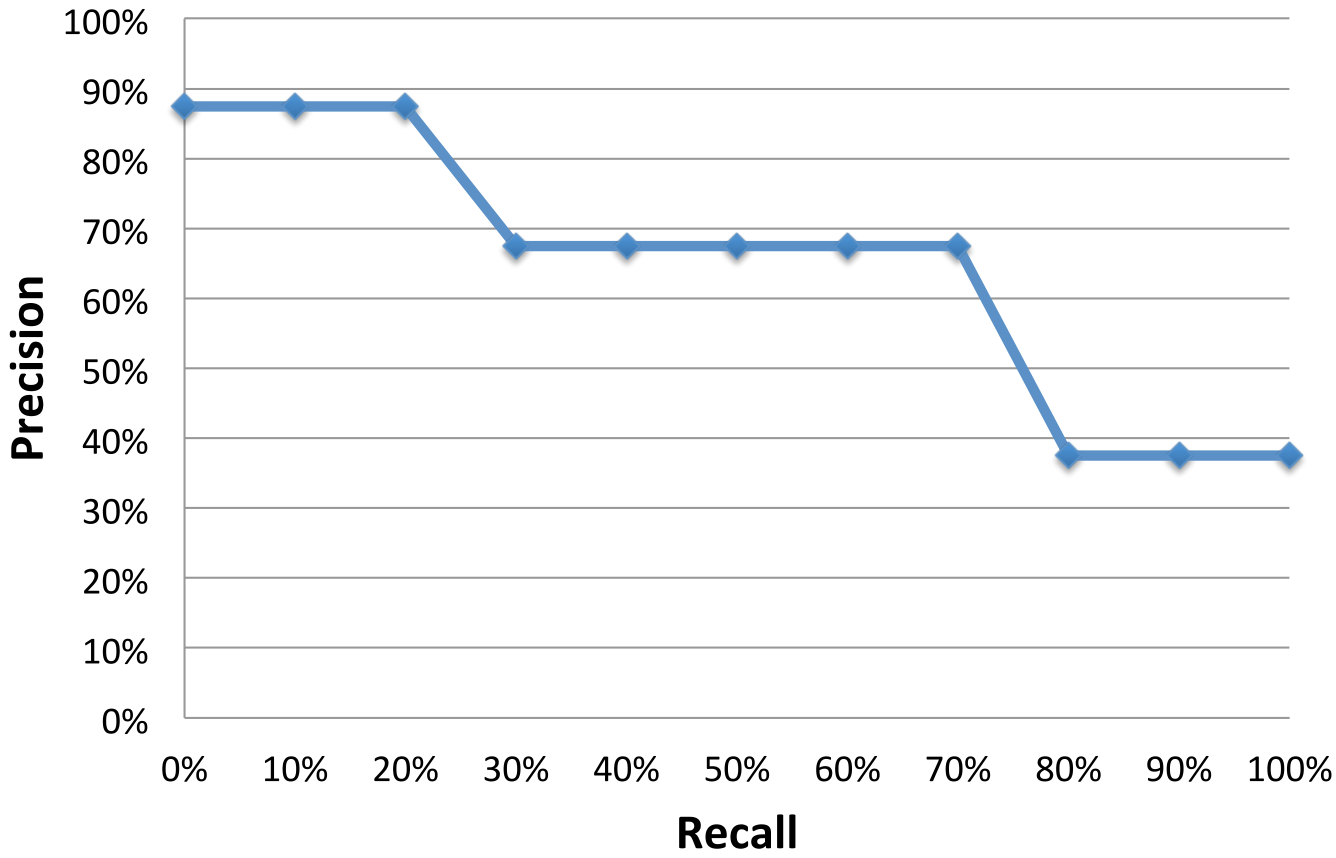
$$\boxed{\overline{P}(r_j) = \sum_{i=1}^{N_q} \frac{P_i(r_j)}{N_q}}$$

- Onde:
  - $\overline{P}(r_j)$ é a média entre as precisões para o nível de revocação r_j
  - $P_i(r_j)$ é a precisão no nível de revocação r_j para a i-ésima consulta

Exemplo – Precisão e Revocação

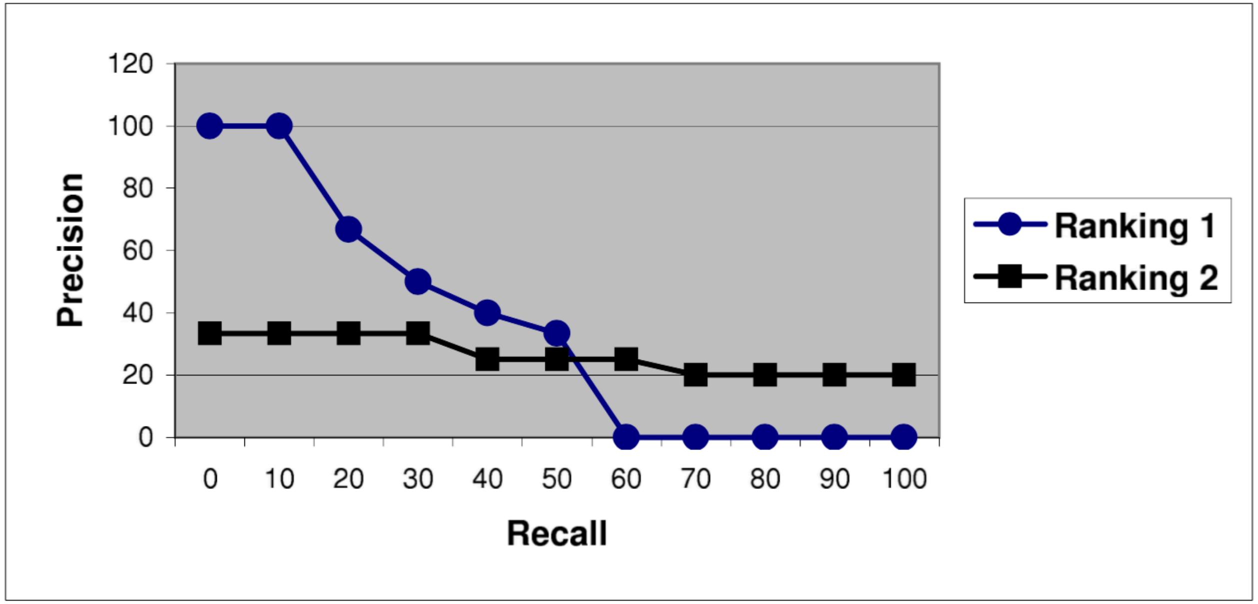
- Para a média das consultas q_1 and q_2

Recall	Precision
0,0	0,88
0,1	0,88
0,2	0,88
0,3	0,68
0,4	0,68
0,5	0,68
0,6	0,68
0,7	0,68
0,8	0,38
0,9	0,38
1,0	0,38



Precisão e Revocação
[Baeza-Yates & Ribeiro-Neto, 2013]

- Curvas de precisão-revocação média são normalmente utilizadas para comparar a *qualidade* de algoritmos de RI distintos



Adequação de Precisão e Revocação
[Baeza-Yates & Ribeiro-Neto, 2013]

- Precisão e revocação têm sido amplamente utilizadas para avaliar a *qualidade* de algoritmos de RI
- Contudo, existem problemas com essas duas medidas, como:
  - A estimativa da revocação máxima para uma consulta requer um conhecimento detalhado de todos os documentos da coleção
  - Em muitas situações, o uso de uma só medida poderia ser mais apropriado

Sumários com um único valor: P@N, MAP, R, F [Baeza-Yates & Ribeiro-Neto, 2013]

- Curvas de precisão-revocação média constituem métricas de avaliação padrão para sistemas de RI
- Existem situações nas quais deseja-se avaliar a *qualidade* da recuperação para consultas individuais
    1. Calcular a média de várias consultas pode encobrir anomalias importantes nos algoritmos de RI estudados
    2. Podemos estar interessados em investigar se um algoritmo é melhor do que outro em todas as consultas
- Nessas situações, um valor único de precisão (para cada consulta) pode ser usado

P@N [Baeza-Yates & Ribeiro-Neto, 2013]

- No caso de máquinas de busca para Web, a maioria das buscas não precisa de alta revocação
- Quanto maior o número de documentos relevantes no topo do *ranking*, mais positiva é a impressão dos usuários
- *Precision at 5* (*P@5*) e *at 10* (*P@10*) medem a precisão quando 5 ou 10 documentos forem vistos
- Essas métricas avaliam se os usuários estão recebendo documentos relevantes no topo ou não

P@5 e *P@10* – Exemplo

[Baeza-Yates & Ribeiro-Neto, 2013]

- Para exemplificar, considere o *ranking* para uma consulta q sendo:

01. d_{123} v	06. d_{9} v	11. d_{38}
02. d_{84}	07. d_{511}	12. d_{48}
03. d_{56} v	08. d_{129}	13. d_{250}
04. d_{6}	09. d_{187}	14. d_{113}
05. d_{8}	10. d_{25} v	15. d_{3} v

  - Para essa consulta, *P@5* = 2/5 = 40% e *P@10* = 4/10 = 40%
  - Além disso, podemos computar a média de *P@5* e *P@10* para umas 100 consultas, por exemplo
  - Essas métricas fornecem uma avaliação da impressão do usuário sobre os resultados (raramente acessam além 2ª página de resultados na Web)

MAP_i [Baeza-Yates & Ribeiro-Neto, 2013]

- Precisão média para a consulta q_i

$$MAP_i = \frac{1}{|R_i|}\sum_{k=1}^{|R_i|} P\left(R_i[k]\right)$$

- Onde:
  - $|R_i|$ é o número de documentos relevantes para a consulta q_i
  - $R_i[k]$ é uma referência ao k-ésimo documento em R_i
  - $P(R_i[k])$ é a precisão quando o documento $R_i[k]$ é observado no *ranking* de q_i
    - Se aquele documento não for recuperado $P(R_i[k])$ assume o valor zero

MAP *(Mean Average Precision)*

[Baeza-Yates & Ribeiro-Neto, 2013]

- Média das precisões médias para um conjunto de consultas

$$\boxed{MAP = \frac{1}{|N_q|}\sum_{i=1}^{|N_q|} MAP_i}$$

- Onde:
  - $|N_q|$ é o número total de consultas

Exemplo – *MAP*

$$MAP_i = \frac{1}{|R_i|}\sum_{k=1}^{|R_i|} P(R_i[k])$$

$R_{q1} = \{d_3, d_5, d_9\}$ e $R_{q2} = \{d_1, d_2, d_6, d_9\}$

Ranking para q_1

#	A
1	d_1
2	d_3 v
3	d_5 v
4	d_9 v
5	d_2

Recall	Precision
0,33	0,5
0,67	0,67
1,0	0,75

MAP_1=1/3*(1/2+2/3+3/4)=**0,6389**

Ranking para q_2

#	A
1	d_9 v
2	d_3
3	d_4
4	d_1 v
5	d_2 v

Recall	Precision
0,25	1,0
0,5	0,5
0,75	0,6

MAP_2=1/4*(1/1+2/4+3/5+0) = **0,525**

$$MAP = \frac{1}{|N_q|}\sum_{i=1}^{|N_q|} MAP_i$$

MAP=1/2*(0,6389+0,525) =**0,5819**

R-Precision

- Seja R o número total de documentos relevantes para uma dada consulta
- A ideia é computar a precisão na R-ésima posição do *ranking*
- A medida *R-precision* é útil para observar o comportamento de um algoritmo para consultas individuais
- Adicionalmente, pode-se computar uma média de *R-precision* sobre um conjunto de consultas
  - Entretanto, usar um único valor para avaliar um algoritmo para várias consultas pode ser bastante impreciso

R-Precision – Exemplo

[Baeza-Yates & Ribeiro-Neto, 2013]

- Para exemplificar, considere o *ranking* para uma consulta q sendo $R_q = \{d_3, d_5, d_9, d_{25}, d_{39}, d_{44}, d_{56}, d_{71}, d_{89}, d_{123}\}$:

01. d_{123} v	06. d_9 v	11. d_{38}
02. d_{84}	07. d_{511}	12. d_{48}
03. d_{56} v	08. d_{129}	13. d_{250}
04. d_6	09. d_{187}	14. d_{113}
05. d_8	10. d_{25} v	15. d_3 v

- Para a consulta q, o valor de R é 10 e existem 4 relevantes dentre os top 10 documentos no *ranking*
- Então, o valor de *R-Precision* para essa consulta é 0.4

F-Measure (Medida F)

- Medida que combina revocação e precisão
- Ideia é permitir ao usuário especificar se ele está mais interessado em revocação ou precisão

$$F_{\beta}(j) = \frac{(1+\beta^{2}) \times P(j) \times r(j)}{(\beta^{2} \times P(j)) + r(j)}$$

- Onde:
  - $r(j)$ é o *recall* na j-ésima posição do *ranking*
  - $P(j)$ é a *precision* na j-ésima posição do *ranking*
  - *ß* é um parâmetro especificado pelo usuário (*ß*≥0). Ex.:
    - *ß* = 1, mesma ênfase à *recall* e *precision* (**F1-measure** → média harmônica)
    - *ß* = 2, *recall* enfatizada 2 vezes em relação à *precision*
    - *ß* = 0.5, *precision* enfatizada 2 vezes em relação à *recall*

F_1-Measure
[Baeza-Yates & Ribeiro-Neto, 2013]

- A função F_1 assume valores no intervalo $[0,1]$
  - Sendo 0 quando nenhum documento tiver sido recuperado e 1 quando todos os documentos ranqueados são relevantes
- Além disso, a média harmônica F_1 assume valores altos somente quando ambos revocação e precisão são altos
- Para maximizar F_1 é requerido encontrar o melhor compromisso possível entre revocação e precisão

nDCG (*Normalized Discounted Cumulative Gain*) [Baeza-Yates & Ribeiro-Neto, 2013]

- Precisão e revocação permitem apenas julgamentos de relevância binários
- Como resultado, não há distinção entre documentos "altamente" relevantes e documentos "moderadamente" relevantes
- Essas limitações podem ser superadas pela adoção de julgamentos de relevância graduais e métricas que os combinem

nDCG (*Normalized Discounted Cumulative Gain*) [Jannach et al., 2010]

- *Discounted cumulative gain (DCG)*
  - Fator de redução logarítmico

$$DCG_{pos} = rel_1 + \sum_{i=2}^{pos} \frac{rel_i}{log_2 i}$$

Onde:
- pos denota a posição até a qual é acumulada a relevância
- rel_i retorna a relevância do documento na posição i
- R é o conjunto de documentos relevantes para uma consulta

- *Idealized discounted cumulative gain (IDCG)*
  - Ranking ideal: assume documentos ordenados em ordem decrescente de relevância

pos_r = **|*R*|**, se |*R*|<*pos*; ***pos***, caso contrário

$$IDCG_{pos} = rel_1 + \sum_{i=2}^{pos\_r} \frac{rel_i}{log_2 i}$$

- ***Normalized discounted cumulative gain (nDCG)***
  - Normalizado no intervalo $[0..1]$

$$\boxed{nDCG_{pos} = \frac{DCG_{pos}}{IDCG_{pos}}}$$

Exemplo – *nDCG*

Escala de julgamento: 0–3 (0 documentos não relevantes, 3 para documentos altamente relevantes)

- Sendo $|R|=3$

$R_{q1} = \{d_3\,(\text{rel}=1),\ d_5\,(\text{rel}=1),\ d_9\,(\text{rel}=1)\}$

Ranking para q_1

#	A
1	d_1
2	d_3 ν
3	d_5 ν
4	d_9 ν
5	d_2

$$nDCG_5 = \frac{DCG_5}{IDCG_5} \cong \frac{2{,}13}{2{,}63} \cong \mathbf{0,81}$$

$$DCG_5 = rel_1 + \sum_{i=2}^{5} \frac{rel_i}{log_2 i}$$
$$= 0 + \frac{1}{log_2 2} + \frac{1}{log_2 3} + \frac{1}{log_2 4} + \frac{0}{log_2 5} \cong 2{,}13$$

$$IDCG_5 = rel_1 + \sum_{i=2}^{3} \frac{rel_i}{log_2 i}$$
$$= 1 + \frac{1}{log_2 2} + \frac{1}{log_2 3} \cong 2{,}63$$

Exemplo – *nDCG*

Escala de julgamento: 0–3 (0 documentos não relevantes, 3 para documentos altamente relevantes)

- Sendo $|R|=3$

$R_{q2} = \{d_4\,(\text{rel}=1),\ d_6\,(\text{rel}=2),\ d_8\,(\text{rel}=3)\}$

Ranking para q_2

#	A
1	d_1
2	d_4 ν
3	d_6 ν
4	d_8 ν
5	d_2

$$nDCG_5 = \frac{DCG_5}{IDCG_5} \cong \frac{3{,}76}{5{,}63} \cong \mathbf{0{,}67}$$

$$DCG_5 = rel_1 + \sum_{i=2}^{5} \frac{rel_i}{log_2 i}$$
$$= 0 + \frac{1}{log_2 2} + \frac{2}{log_2 3} + \frac{3}{log_2 4} + \frac{0}{log_2 5} \cong 3{,}76$$

$$IDCG_5 = rel_1 + \sum_{i=2}^{3} \frac{rel_i}{log_2 i}$$
$$= 3 + \frac{2}{log_2 2} + \frac{1}{log_2 3} \cong 5{,}63$$

nDCG [Baeza-Yates & Ribeiro-Neto, 2013]

- Para computar o *nDCG* sobre um conjunto de consultas de teste, precisamos calcular a razão entre a média sobre todas as consultas de *DCG* e *IDCG*
- Dado um conjunto de N_q consultas, o *nDCG* é calculado da seguinte forma:

$$\overline{DCG}[i] = \sum_{j=1}^{N_q} \frac{DCG_j[i]}{N_q} \qquad \overline{IDCG}[i] = \sum_{j=1}^{N_q} \frac{IDCG_j[i]}{N_q}$$

$$\boxed{NDCG[i] = \frac{\overline{DCG}[i]}{\overline{IDCG}[i]}}$$

  - Onde: $[i]$ é equivalente ao pos (posição no *ranking*) nas equações apresentadas anteriormente

Exemplo – *nDCG*

$$\overline{DCG}[i] = \sum_{j=1}^{N_q} \frac{DCG_j[i]}{N_q}$$

$DCG_1[5]=2,13$

$DCG_2[5]=3,76$

$\overline{DCG}[5]=(2,13+3,76)/2$
$=\mathbf{2,95}$

$$\overline{IDCG}[i] = \sum_{j=1}^{N_q} \frac{IDCG_j[i]}{N_q}$$

$IDCG_1[5]=2,63$

$IDCG_2[5]=5,63$

$\overline{IDCG}[5]=(2,63+5,63)/2$
$=\mathbf{4,13}$

$$\boxed{NDCG[i] = \frac{\overline{DCG}[i]}{\overline{IDCG}[i]}}$$

NDCG[5]=2,95/4,13=**0,71**

Discussões sobre *nDCG*

[Baeza-Yates & Ribeiro-Neto, 2013]

- Objetiva levar em consideração múltiplos níveis de julgamentos de relevância
- Tem a vantagem de distinguir documentos “altamente” relevantes de outros “moderadamente” relevantes
- A desvantagem inerente são os múltiplos níveis de julgamento de relevância que são mais difíceis e mais demorados de serem obtidos

Medidas de correlação de *ranking*

[Baeza-Yates & Ribeiro-Neto, 2013]

- Há situações em que
  - Não queremos medir diretamente a relevância
  - Estamos mais interessados em determinar o quão diferentemente uma função de *ranking* varia de uma segunda que é bem conhecida
- Nesses casos, estamos interessados em comparar a ordem relativa produzida por dois *rankings*
- Isto pode ser alcançado através da utilização de funções estatísticas chamadas ***rank correlation metrics***

Medidas de correlação de *ranking*

[Baeza-Yates & Ribeiro-Neto, 2013]

- Sejam os rankings $\mathcal{R}_1$ e $\mathcal{R}_2$
- Uma medida de correlação de *ranking* produz um coeficiente de correlação $\mathcal{C}(\mathcal{R}_1, \mathcal{R}_2)$ com as seguintes propriedades:
  - $-1 \leq \mathcal{C}(\mathcal{R}_1, \mathcal{R}_2) \leq 1$
  - se $\mathcal{C}(\mathcal{R}_1, \mathcal{R}_2) = 1$, a concordância entre os dois *rankings* é perfeita, isto é, eles são o mesmo
  - se $\mathcal{C}(\mathcal{R}_1, \mathcal{R}_2) = -1$, a discordância entre os dois *rankings* é perfeita, eles são o inverso um do outro
  - se $(\mathcal{R}_1, \mathcal{R}_2) = 0$, os dois *rankings* são completamente independentes
  - O aumento dos valores de $\mathcal{C}(\mathcal{R}_1, \mathcal{R}_2)$ implica no aumento de concordância entre os dois *rankings*

Coeficiente de Spearman
[Baeza-Yates & Ribeiro-Neto, 2013]

- O coeficiente de Spearman é provavelmente a medida de correlação de *ranking* mais utilizada
- É baseado em diferenças entre as posições de um mesmo documento em dois rankings
- Seja
  - $s_{1,j}$ a posição do documento d_j no *ranking* $\mathcal{R}_1$
  - $s_{2,j}$ a posição do documento d_j no *ranking* $\mathcal{R}_2$

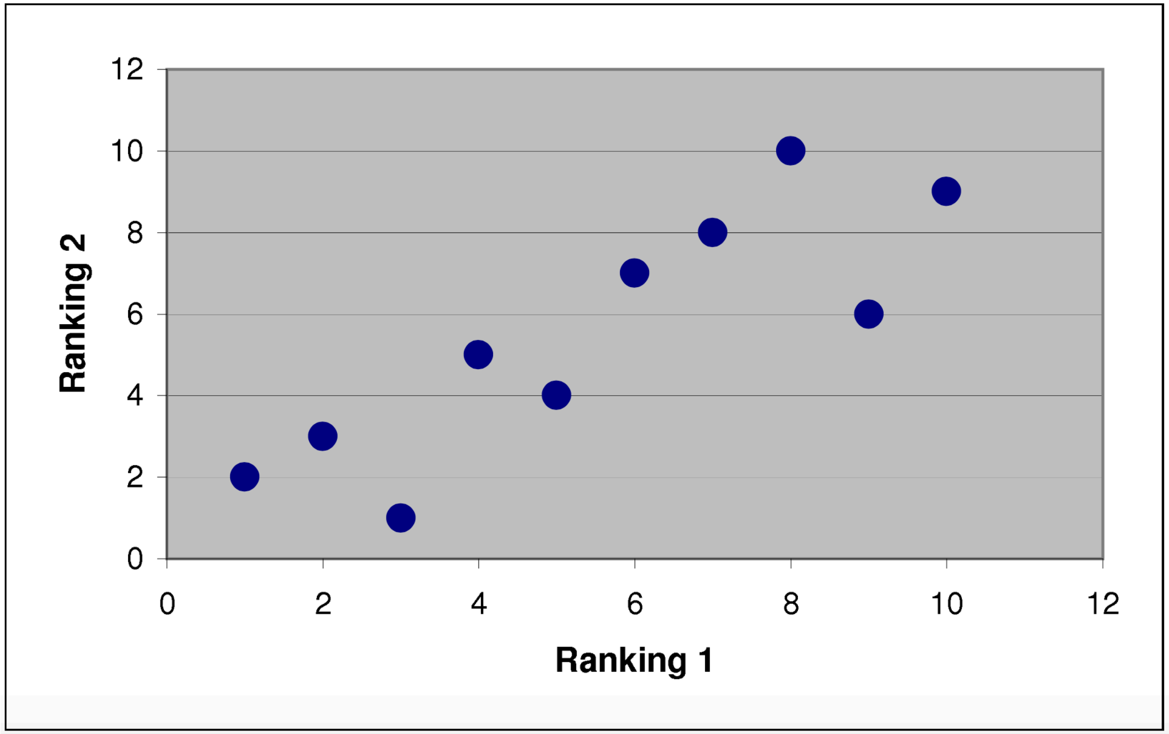
Exemplo – Coeficiente de Spearman
[Baeza-Yates & Ribeiro-Neto, 2013]

- Considere 10 documentos recuperados por dois *rankings* distintos $\mathcal{R}_1$ e $\mathcal{R}_2$

documents	$s_{1,j}$	$s_{2,j}$	$s_{1,j} - s_{2,j}$	$(s_{1,j} - s_{2,j})^2$
d_{123}	1	2	-1	1
d_{84}	2	3	-1	1
d_{56}	3	1	+2	4
d_{6}	4	5	-1	1
d_{8}	5	4	+1	1
d_{9}	6	7	-1	1
d_{511}	7	8	-1	1
d_{129}	8	10	-2	4
d_{187}	9	6	+3	9
d_{25}	10	9	+1	1
Sum of Square Distances				24

Coeficiente de Spearman
[Baeza-Yates & Ribeiro-Neto, 2013]

- Plotando as posições nos rankings para $\mathcal{R}_1$ e $\mathcal{R}_2$ em um sistema de coordenadas bi-dimensional, podemos observar que existe uma forte correlação entre os dois *rankings*



Coeficiente de Spearman
[Baeza-Yates & Ribeiro-Neto, 2013]

- Para produzir uma avaliação quantitativa dessa correlação, somamos os quadrados das diferenças para cada par dos *rankings*
- Se existem K documentos ranqueados, o valor máximo para a soma dos quadrados das diferenças dos *rankings* é dado por

$$K \times (K^2 - 1) / 3$$

- Seja K=10
  - Se os dois *rankings* estão em perfeita discordância, então esse valor é $(10 \times (10^2 - 1))/3$, ou 330
  - Por outro lado, se existe uma concordância completa a soma é 0 (zero)

Coeficiente de Spearman
[Baeza-Yates & Ribeiro-Neto, 2013]

- Consideramos a fração

$$\frac{\sum_{j=1}^{K}(s_{1,j} - s_{2,j})^2}{\frac{K \times (K^2 - 1)}{3}}$$

- Seu valor é
  - 0 quando os dois *rankings* estão em perfeita concordância
  - $+1$ quando eles estão em perfeita discordância
- Se multiplicarmos a fração por 2, seu valor é alterado para o intervalo $[0, +2]$
- A seguir, se subtrairmos o resultado de 1, o valor resultante é alterado para o intervalo $[-1, +1]$

Coeficiente de Spearman

[Baeza-Yates & Ribeiro-Neto, 2013]

- Esse raciocínio sugere a definição da correlação entre dois *rankings* como segue
- Seja $s_{1,j}$ e $s_{2,j}$ as posições de um documento d_j em dois *rankings* $\mathcal{R}_1$ e $\mathcal{R}_2$, respectivamente
- Definimos

$$\boxed{S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times \sum_{j=1}^{K} (s_{1,j} - s_{2,j})^2}{K \times (K^2 - 1)}}$$

- Onde:
  - $S(\mathcal{R}_1, \mathcal{R}_2)$ é o *Spearman rank correlation coefficient*
  - K indica o tamanho dos conjuntos ranqueados

Exemplo – Coeficiente de Spearman

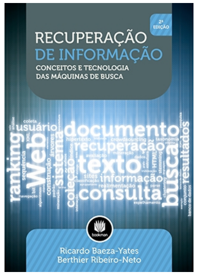
[Baeza-Yates & Ribeiro-Neto, 2013]

- Para os *rankings* do exemplo, temos

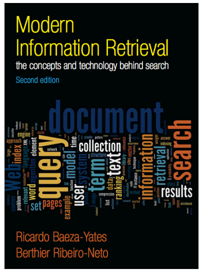
$$S(\mathcal{R}_1, \mathcal{R}_2) = 1 - \frac{6 \times 24}{10 \times (10^2 - 1)} = 1 - \frac{144}{990} = 0.854$$

documents	$s_{1,j}$	$s_{2,j}$	$s_{1,j} - s_{2,j}$	$(s_{1,j} - s_{2,j})^2$
d_{123}	1	2	-1	1
d_{84}	2	3	-1	1
d_{56}	3	1	+2	4
d_6	4	5	-1	1
d_8	5	4	+1	1
d_9	6	7	-1	1
d_{511}	7	8	-1	1
d_{129}	8	10	-2	4
d_{187}	9	6	+3	9
d_{25}	10	9	+1	1
Sum of Square Distances				24

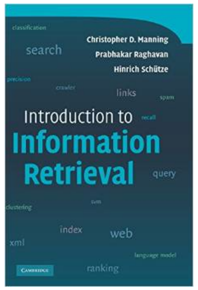
Referências



- Baeza-Yates, R.; Ribeiro-Neto, B. Recuperação de Informação: Conceitos e Tecnologia das Máquinas de Busca. 2 ed. Bookman, 2013.



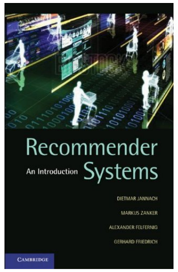
- Baeza-Yates, R.; Ribeiro-Neto, B. Modern Information Retrieval. Wokingham, UK: Addison-Wesley, 2 ed., 2011.



- Manning, C. D.; Raghavan, P.; Schütze, H. Introduction to Information Retrieval. Cambridge University Press, 2008.

  Online edition 2009: http://nlp.stanford.edu/IR-book/

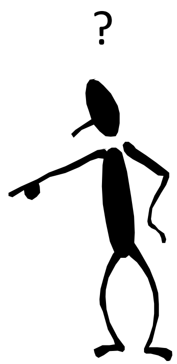
Referências



- Jannach, D.; Zanker, M.; Felfernig, A.; Friedrich, G. Recommender Systems: An Introduction. 1 ed. Cambridge University Press, 2010.

Recuperação da Informação (MAB605)

Dúvidas?



Profa. Giseli Rabello Lopes
giseli@dcc.ufrj.br
CCMN - DCC - Sala E-2012