

Recuperação de Informação

Profa. Patrícia Aparecida Proença Avila
patricia.proenca@ifmg.edu.br

Atividade - aula 06/07

- ▶ Aplique o modelo probabilístico no exercício 1 da lista II.
- ▶ **Lista II:**
- ▶ 1 - Use a coleção de documentos do exercício 6 da lista 1 para resolver as solicitações abaixo:
 - ▶ homem moço;
 - ▶ dizer ir tempo;
 - ▶ dia senhora casa;

Atividade - aula 06/07

► **Lista I - documentos:**

- D1 = {homem estar tempo coisa dizer ir ter}
- D2 = {senhora estar dia moço moço senhora}
- D3 = {senhora vez senhora senhora tempo dizer filho}
- D4 = {casa ir ir dizer ter olho}
- D5 = {olho dia vez dia homem moço tempo}

	d1	d2	d3	d4	d5
casa	0	0	0	1	0
coisa	1	0	0	0	0
dia	0	1	0	0	2
dizer	1	0	1	1	0
estar	1	1	0	0	0
filho	0	0	1	0	0
homem	1	0	0	0	1
ir	1	0	0	2	0
moço	0	2	0	0	1
olho	0	0	0	1	1
senhora	0	2	3	0	0
tempo	1	0	1	0	1
ter	1	0	0	1	0
vez	0	0	1	0	1

Não conhecemos o conjunto de relevantes: $R = r_i = 0$

$$\text{sim}(d_j, q) \sim \sum_{k_i \in q \wedge k_i \in d_j} \log \left(\frac{N + 0,5}{n_i + 0,5} \right)$$

Atividade - aula 06/07

► Homem moço

► D1: $\log ((5+0.5)/(2+0.5)) = 1.137$

► D2: $\log ((5+0,5)/(2+0.5))= 1.137$

► D3: 0

► D4: 0

► D5: $\log ((5+0.5)/(2+0.5)) + \log (((5+0.5)/(5+0.5)))= 2,275$

	d1	d2	d3	d4	d5
casa	0	0	0	1	0
coisa	1	0	0	0	0
dia	0	1	0	0	2
dizer	1	0	1	1	0
estar	1	1	0	0	0
filho	0	0	1	0	0
homem	1	0	0	0	1
ir	1	0	0	2	0
moço	0	2	0	0	1
olho	0	0	0	1	1
senhora	0	2	3	0	0
tempo	1	0	1	0	1
ter	1	0	0	1	0
vez	0	0	1	0	1

$$sim(d_j, q) \sim \sum_{k_i \in q \wedge k_i \in d_j} \log \left(\frac{N + 0,5}{n_i + 0,5} \right)$$

Atividade - aula 06/07

► dizer ir tempo

► D1: $\log((5+0.5)/(3+0.5)) + \log((5+0.5)/(2+0.5)) + \log((5+0.5)/(3+0.5)) = 2,441$

► D2: 0

► D3: $\log((5+0.5)/(3+0.5)) + \log((5+0.5)/(3+0.5)) = 1,304$

► D4: $\log((5+0.5)/(3+0.5)) + \log((5+0.5)/(2+0.5)) = 1,789$

► D5: $\log((5+0.5)/(3+0.5)) = 0,652$

	d1	d2	d3	d4	d5
casa	0	0	0	1	0
coisa	1	0	0	0	0
dia	0	1	0	0	2
dizer	1	0	1	1	0
estar	1	1	0	0	0
filho	0	0	1	0	0
homem	1	0	0	0	1
ir	1	0	0	2	0
moço	0	2	0	0	1
olho	0	0	0	1	1
senhora	0	2	3	0	0
tempo	1	0	1	0	1
ter	1	0	0	1	0
vez	0	0	1	0	1

$$\text{sim}(d_j, q) \sim \sum_{k_i \in q \wedge k_i \in d_j} \log \left(\frac{N + 0,5}{n_i + 0,5} \right)$$

Atividade - aula 06/07

- ▶ Dia senhora casa
- ▶ D1: 0
- ▶ D2: $\log((5+0.5)/(2+0.5)) + \log((5+0.5)/(2+0.5)) = 2.275$
- ▶ D3: $\log((5+0.5)/(2+0.5)) = 1.137$
- ▶ D4: $\log((5+0.5)/(1+0.5)) = 1.874$
- ▶ D5: $\log((5+0.5)/(1+0.5)) = 1.874$

	d1	d2	d3	d4	d5
casa	0	0	0	1	0
coisa	1	0	0	0	0
dia	0	1	0	0	2
dizer	1	0	1	1	0
estar	1	1	0	0	0
filho	0	0	1	0	0
homem	1	0	0	0	1
ir	1	0	0	2	0
moço	0	2	0	0	1
olho	0	0	0	1	1
senhora	0	2	3	0	0
tempo	1	0	1	0	1
ter	1	0	0	1	0
vez	0	0	1	0	1

$$\text{sim}(d_j, q) \sim \sum_{k_i \in q \wedge k_i \in d_j} \log \left(\frac{N + 0,5}{n_i + 0,5} \right)$$

Atividade Aula - 19/07

- ▶ Considere uma coleção de referência e um conjunto de consultas para teste. Suponha que os conjuntos $R1$, $R2$ e $R3$ de documentos relevantes para as consultas $q1$, $q2$ e $q3$, respectivamente, tenham sido determinados por um grupo de especialistas.
 - ▶ $R1 = \{d3, d7, d12, d13, d26, d68\}$
 - ▶ $R2 = \{d1, d2, d9, d24, d51, d52, d70, d82\}$
 - ▶ $R3 = \{d2, d3, d6, d16, d20\}$
- ▶ Considere que um novo algoritmo de recuperação chamado XYZ foi recém projetado. Suponha que esse algoritmo retorne, para as consultas $q1$, $q2$ e $q3$, os seguintes rankings de documentos (primeiras quinze posições):
 - ▶ Consulta $q1$ (algoritmo XYZ) = $\{d1, d9, d26, d15, d2, d10, d74, d68, d32, d3, d53, d39, d56, d11, d4\}$.
 - ▶ Consulta $q2$ (algoritmo XYZ) = $\{d3, d7, d8, d9, d19, d16, d37, d24, d20, d80, d67, d50, d46, d51, d29\}$.
 - ▶ Consulta $q3$ (algoritmo XYZ) = $\{d2, d30, d25, d3, d9, d7d6, d39, d75, d19, d26 d16, d20, d51, d1\}$.

Atividade Aula - 19/07

- ▶ a) Calcule os níveis de precisão e revocação para cada uma das consultas.
- ▶ $R1 = \{d3, d7, d12, d13, d26, d68\}$
- ▶ Consulta q1 (algoritmo XYZ) = $\{d1, d9, \underline{d26}, d15, d2, d10, d74, \underline{d68}, d32, \underline{d3}, d53, d39, d56, d11, d4\}$.
- ▶ $p = 1/3 = 33,3\%$ $r = 1/6 = 16,6\%$
- ▶ $p = 2/8 = 25\%$ $r = 2/6 = 33,3\%$
- ▶ $p = 3/10 = 30\%$ $r = 3/6 = 20\%$

Revocação	Precisão
0	33,3
10	33,3
20	30
30	25
40	0
50	0
60	0
70	0
80	0
90	0
100	0

Atividade Aula - 19/07

- ▶ a) Calcule os níveis de precisão e revocação para cada uma das consultas.
- ▶ $R2 = \{d1, d2, d9, d24, d51, d52, d70, d82\}$
- ▶ Consulta q2 (algoritmo XYZ) = $\{d3, d7, d8, \underline{d9}, d19, d16, d37, \underline{d24}, d20, d80, d67, d50, d46, \underline{d51}, d29\}$.
- ▶ $p = 1/4 = 25\%$ $r = 1/8 = 12,5\%$
- ▶ $p = 2/8 = 25\%$ $r = 2/8 = 25\%$
- ▶ $p = 3/14 = 21,4\%$ $r = 3/8 = 37,5\%$

Revocação	Precisão
0	25
10	25
20	25
30	21,4
40	0
50	0
60	0
70	0
80	0
90	0
100	0

Atividade Aula - 19/07

- ▶ a) Calcule os níveis de precisão e revocação para cada uma das consultas.
- ▶ $R3 = \{d2, d3, d6, d16, d20\}$
- ▶ Consulta q3 (algoritmo XYZ) = $\{\underline{d2}, d30, d25, \underline{d3}, d9, d7, \underline{d6}, d39, d75, d19, d26, \underline{d16}, \underline{d20}, d51, d1\}$.
- ▶ $p = 1/1 = 100\% \quad r = 1/5 = 20\%$
- ▶ $p = 2/4 = 50\% \quad r = 2/5 = 40\%$
- ▶ $p = 3/7 = 42,8\% \quad r = 3/5 = 60\%$
- ▶ $p = 4/12 = 33,3\% \quad r = 4/5 = 80\%$
- ▶ $p = 5/13 = 38\% \quad r = 5/5 = 100\%$

Revocação	Precisão
0	100%
10	100%
20	100%
30	50%
40	50%
50	42,8
60	42,8
70	33,3
80	33,3
90	38
100	38

Atividade Aula - 19/07

▶ **PARA VOCÊS PENSAREM!!!!!!**

- ▶ b) Construa o gráfico de precisão versus revocação para cada uma das consultas.
- ▶ c) Encontre a precisão média do algoritmo XYZ e faça o gráfico dos valores médio de precisão versus revocação.

Atividade Aula - 20/07

- ▶ Com base na coleção de documentos apresentada na exercício da aula anterior, responda:
- ▶ a) Os valores de precisão $P@n$ nas posições 5, 10 e 15 para cada uma das três consultas.
- ▶ *Consulta q1 (algoritmo XYZ) = {d1, d9, d26, d15, d2, d10, d74, d68, d32, d3, d53, d39, d56, d11, d4}.*
 - ▶ $P@5 = 1/5 = 20\%$ $P@10 = 3/10 = 30\%$ $P@15 = 3/15 = 20\%$
- ▶ *Consulta q2 (algoritmo XYZ) = {d3, d7, d8, d9, d19, d16, d37, d24, d20, d80, d67, d50, d46, d51, d29}.*
 - ▶ $P@5 = 1/5 = 20\%$ $P@10 = 2/10 = 20\%$ $P@15 = 3/15 = 20\%$
- ▶ *Consulta q3 (algoritmo XYZ) = {d2, d30, d25, d3, d9, d7, d6, d39, d75, d19, d26, d16, d20, d51, d1}.*
 - ▶ $P@5 = 2/5 = 40\%$ $P@10 = 3/10 = 30\%$ $P@15 = 5/15 = 33,3\%$
- ▶ b) Para qual consulta a impressão dos usuários será mais positiva? Justifique.

Ame hoje.
♥ Perdoe hoje. ♥
Demonstre hoje.
Não deixe nada
para amanhã.
Somos instantes
e num instante
não somos nada.