

Avaliação da Recuperação

Wendel Melo

Faculdade de Computação
Universidade Federal de Uberlândia

Recuperação da Informação

Adaptado do Material da Prof^a Vanessa Braganholo - IC/UFF

Avaliação de sistemas de busca

- Podemos avaliar um sistema de busca quanto a uma série de fatores:

Avaliação de sistemas de busca

- Podemos avaliar um sistema de busca quanto a uma série de fatores:
 - Desempenho;
 - Escalabilidade;
 - Utilização de recursos;
 - Interfaces de entrada e saída;
 - ...

Avaliação de sistemas de busca

- Podemos avaliar um sistema de busca quanto a uma série de fatores:
 - Desempenho;
 - Escalabilidade;
 - Utilização de recursos;
 - Interfaces de entrada e saída;
 - ...
- No entanto, os termos *avaliação de sistemas de busca* e/ou *avaliação de recuperação de informação* são utilizados no contexto de aferir a qualidade da resposta de sistemas de busca.

Avaliação da Recuperação da Informação

- É um processo sistemático no qual se associa uma métrica quantitativa aos resultados produzidos por um sistema de RI (busca) em resposta a um conjunto de consultas de usuário;
- A métrica associada deve estar diretamente associada à relevância dos resultados para os usuários;
- Assim, embora a métrica seja quantitativa, o objetivo final é avaliar a qualidade do sistema com respeito às respostas fornecidas.

Para que avaliar um sistema de RI?

Para que avaliar um sistema de RI?

- Para saber se o mesmo está cumprindo seu papel e como está desempenhando;

Para que avaliar um sistema de RI?

- Para saber se o mesmo está cumprindo seu papel e como está desempenhando;
- Para comparar com outros sistemas de RI;

Para que avaliar um sistema de RI?

- Para saber se o mesmo está cumprindo seu papel e como está desempenhando;
- Para comparar com outros sistemas de RI;
- Para avaliar se modificações no modelo de RI e/ou no ranqueamento trazem melhorias ao sistema;

Para que avaliar um sistema de RI?

- Para saber se o mesmo está cumprindo seu papel e como está desempenhando;
- Para comparar com outros sistemas de RI;
- Para avaliar se modificações no modelo de RI e/ou no ranqueamento trazem melhorias ao sistema;
- Para saber com quais tipos de base de dados/consultas o sistema/modelo pode funcionar melhor.

Avaliação da RI

- Na prática, pode ser difícil avaliar um sistema de RI, pois a relevância pode depender de muitos fatores subjetivos para cada usuário;
- Ainda assim, costuma-se associar métricas aos resultados da consulta por:
 - Simplicidade;
 - Poder repetir experimentos diversas vezes;
 - Custo relativamente baixo.

Avaliação da RI

Qual seria a metodologia mais intuitiva para a avaliação de sistemas de RI?

Avaliação da RI

Qual seria a metodologia mais intuitiva para a avaliação de sistemas de RI?

- Convocar especialistas para analisar a resposta produzida pelo sistema de RI?

Avaliação da RI

Qual seria a metodologia mais intuitiva para a avaliação de sistemas de RI?

- Convocar especialistas para analisar a resposta produzida pelo sistema de RI?
 - Seria custoso, demorado e não prático para avaliar diversas configurações distintas.

Avaliação da RI

Qual seria a metodologia mais intuitiva para a avaliação de sistemas de RI?

- ~~Convocar especialistas para analisar a resposta produzida pelo sistema de RI?~~
 - ~~Seria custoso, demorado e não prático para avaliar diversas configurações distintas.~~
- Comparar o resultado produzido por um sistema de RI com o produzido por especialistas humanos;

Avaliação da RI

Qual seria a metodologia mais intuitiva para a avaliação de sistemas de RI?

- Comparar o resultado produzido por um sistema de RI com o produzido por especialistas humanos;
- É comum então a adoção das chamadas **coleções de referência**.
- Coleções de referência: bases de documentos de referência onde especialistas já apontaram quais seriam os documentos relevantes para determinadas consultas específicas.

Exemplos de coleções de referência

- TREC (Text Retrieval Conference)
- CF (Cystic Fibrosis Database) na MEDLINE
- <http://www.search-engines-book.com/collections/>
 - CACM (Communications of the ACM)
 - Wikipedia

Métricas de Recuperação

- Dada uma coleção de referência, com uma conhecida requisição de informação (consulta) I , sejam:
 - R : conjunto de documentos relevantes (apontados por especialistas);
 - A : conjunto resposta do algoritmo de RI sendo avaliado.
- A partir desses dois conjuntos, definimos as métricas de precisão e revocação (cobertura).

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão:** fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão:** fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

- A precisão se remete ao quanto os resultados da busca são uteis.

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão**: fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

- A precisão se remete ao quanto os resultados da busca são uteis.
- A precisão por si só é uma medida completa para avaliação?

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão:** fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

- A precisão se remete ao quanto os resultados da busca são uteis.
- A precisão por si só é uma medida completa para avaliação?
- E se houverem 100 docs relevantes e o algoritmo recuperar 10 docs, sendo 9 relevantes?

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão**: fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

- A precisão se remete ao quanto os resultados da busca são uteis.
- A precisão por si só é uma medida completa para avaliação?
- E se houverem 100 docs relevantes e o algoritmo recuperar 10 docs, sendo 9 relevantes?
- A precisão será de 90%! Mas só se recuperou 9 de 100 relevantes...

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão:** fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

Métricas de Recuperação

- R : conjunto de documentos relevantes (apontado por especialistas);
- A : conjunto resposta do algoritmo de RI sendo avaliado.
- **Precisão**: fração dos documentos recuperados que é relevante:

$$precisão = p = \frac{|R \cap A|}{|A|}$$

- **Revocação (cobertura)**: fração dos documentos relevantes que é recuperada

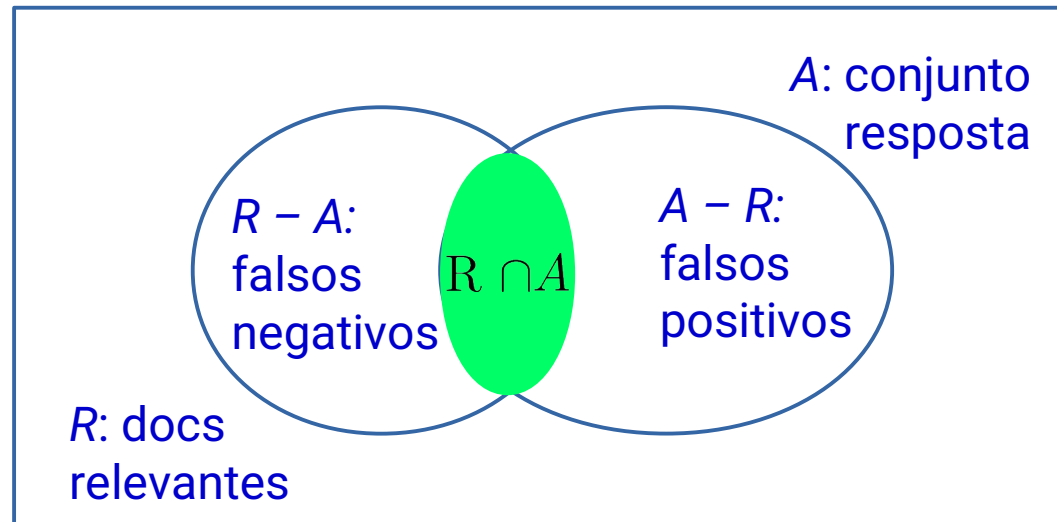
$$revocação = r = \frac{|R \cap A|}{|R|}$$

- A revocação se refere a quão completos os resultados estão

Precisão e Revocação

- Pode-se obter 100% de revocação se todos os documentos forem retornados em todas as consultas!
 - Todavia, a precisão será baixa

Precisão e Revocação



- A precisão é máxima quando o conjunto $A-R$ (falsos positivos) é vazio;
- A revocação é máxima quando o conjunto $R-A$ (falsos negativos) é vazio

Precisão e Revocação

- Observe que as métricas de precisão e revocação supõem que todos os documentos do conjunto resposta foram avaliados;
- Na prática, muitas vezes o usuário examina uma parte da resposta de acordo com a ordem de ranqueamento;
- Assim, as medidas de precisão e revocação variam conforme o usuário vai avaliando as respostas, e costuma-se plotar uma curva de precisão versus revocação.

Precisão e Revocação

- Suponha que, para uma consulta q de uma coleção de referência, o conjunto R tenha 10 docs relevantes e que os docs recuperados pelo algoritmo, já com ranqueamento, foram:

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

- Quando o usuário olhar apenas o 1º doc (d_{12}), pelo fato do mesmo ser relevante, teremos 10% de revocação (1 doc relevante observado de um total 10) e 100% de precisão (todos os docs vistos até então são relevantes).

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

- O próximo documento relevante é o terceiro (d_{56});
- Nesse ponto, teremos 20% de revocação (dois docs relevantes observados de um total de 10) e 67% de precisão (dos três docs vistos até então, dois são relevantes).

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

- O próximo documento relevante é o sexto (d_9);
- Nesse ponto, teremos 30% de revocação (três docs relevantes observados de um total de 10) e 50% de precisão (dos seis docs vistos até então, três são relevantes).

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com
* também pertencem a R , isto é,
são relevantes

- O próximo documento relevante é o décimo (d_{25});
- Nesse ponto, teremos 40% de revocação (quatro docs relevantes observados de um total de 10) e 40% de precisão (dos dez docs vistos ate então, quatro são relevantes).

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

- Por fim, o último documento relevante é o décimo quinto (d_3);
- Nesse ponto, teremos 50% de revocação (5 docs relevantes observados de um total de 10) e 33% de precisão (dos 15 docs vistos ate então, 5 são relevantes).

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

Assim, montamos a seguinte tabela:

| Revocação | Precisão |
|-----------|----------|
| 10% | 100% |
| 20% | 67% |
| 30% | 50% |
| 40% | 40% |
| 50% | 33% |

Só é necessário tabelar nas revocações relativas a documentos relevantes!

Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

Assim, montamos a seguinte tabela:

| Revocação | Precisão |
|-----------|----------|
| 10% | 100% |
| 20% | 67% |
| 30% | 50% |
| 40% | 40% |
| 50% | 33% |

Montamos então um gráfico com a precisão em cada nível de revocação.

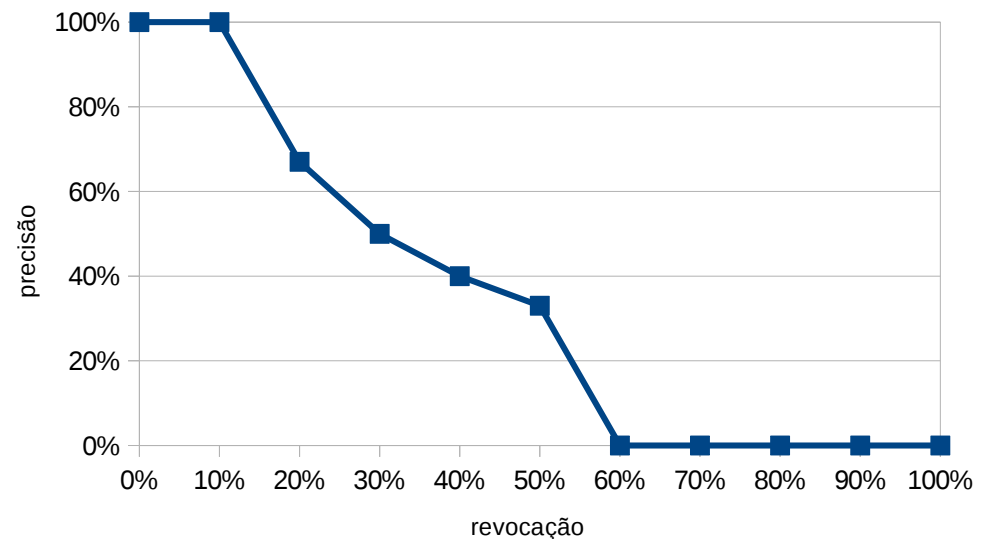
Precisão e Revocação

| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

Assim, montamos a seguinte tabela:

| Revocação | Precisão |
|-----------|----------|
| 10% | 100% |
| 20% | 67% |
| 30% | 50% |
| 40% | 40% |
| 50% | 33% |



Precisão e Revocação

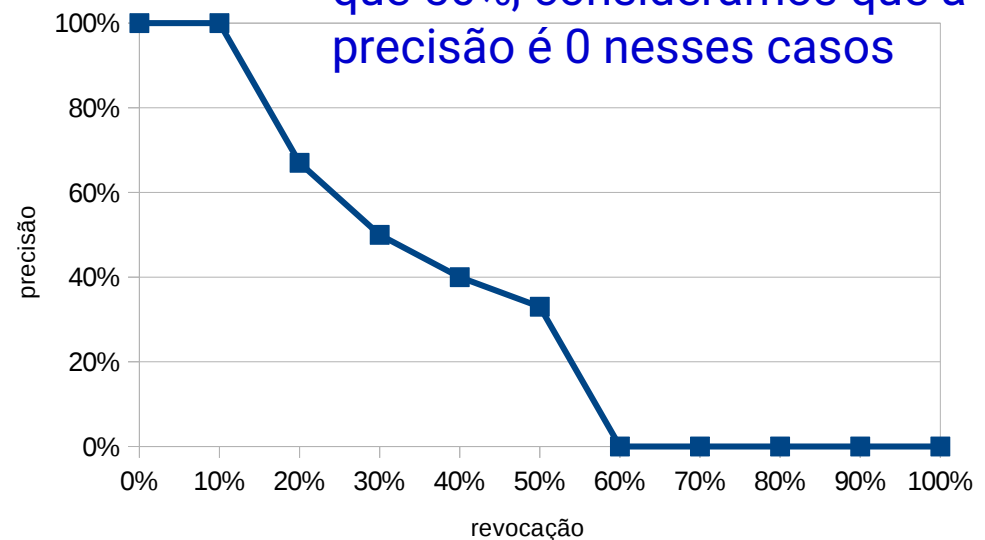
| | | | |
|----------------|--------------|-----------------|---------------|
| 1 - d_{12}^* | 5 - d_8 | 9 - d_{18} | 13 - d_{27} |
| 2 - d_{84} | 6 - d_9^* | 10 - d_{25}^* | 14 - d_{11} |
| 3 - d_{56}^* | 7 - d_{51} | 11 - d_{38} | 15 - d_3^* |
| 4 - d_6 | 8 - d_{19} | 12 - d_{48} | |

Documentos com * também pertencem a R , isto é, são relevantes

Assim, montamos a seguinte tabela:

| Revocação | Precisão |
|-----------|----------|
| 10% | 100% |
| 20% | 67% |
| 30% | 50% |
| 40% | 40% |
| 50% | 33% |

Como não há revocação maior que 50%, consideramos que a precisão é 0 nesses casos



Precisão e Revocação

- Em geral, adotamos 11 níveis padrão de revocação:

$$r_0 = 0\%, \quad r_1 = 10\%, \quad r_2 = 20\%, \quad \dots \quad r_{10} = 100\%$$

- Em outras palavras, só plotamos no gráfico as precisões relativas às revocações de 0%, 10%, 20%, ..., 90%, 100%
- Para evitar picos nos gráficos, adota-se a seguinte regra de interpolação:

$$p(r_j) = \text{maior precisão nas revocações maiores ou iguais a } r_j$$

Regra de interpolação

Precisão e Revocação

- Para evitar picos nos gráficos, consideramos que a precisão $p(r_j)$ na revocação r_j é dada por:

$$p(r_j) = \text{maior precisão nas revocações maiores ou iguais a } r_j$$

- Ou, equivalentemente:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

- Assim, teremos uma curva não crescente, pois, em cada ponto, considera-se a precisão como o maior valor tabelado daquele ponto em diante.

Precisão e Revocação

- **Exemplo:** Suponha que, para uma consulta q_2 conhecida, existam 4 docs relevantes, e que a resposta do algoritmo avaliado foi:

| | |
|--------------|----------------|
| 1 - d_7 | 5 - d_{79} |
| 2 - d_6^* | 6 - d_{30} |
| 3 - d_2 | 7 - d_1^* |
| 4 - d_{13} | 8 - d_{15}^* |

Documentos com * também pertencem a R , isto é, são relevantes

- O segundo resultado é o primeiro relevante. Assim, para a revocação de 25%, temos $1/2 = 50\%$ de precisão.
- O sétimo resultado é o segundo relevante. Assim, para revocação de 50%, temos $2/7 = 29\%$ de precisão.
- O oitavo resultado é o terceiro relevante. Assim, para a revocação de 75%, temos $3/8 = 38\%$ de precisão.

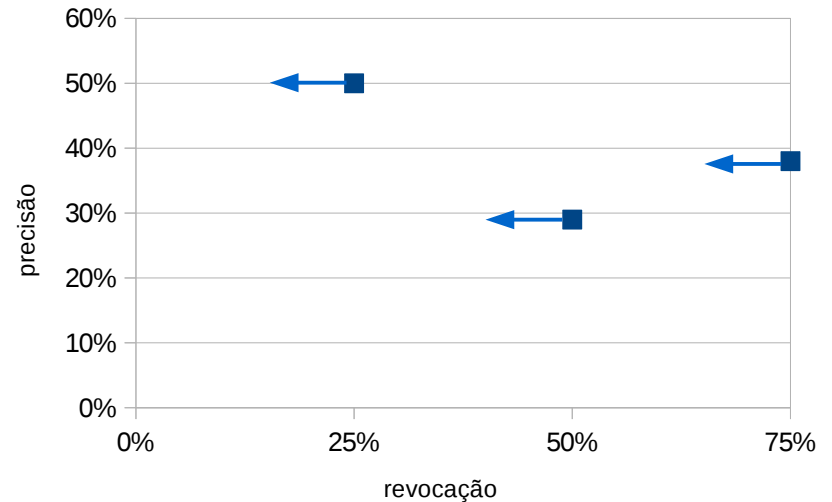
Precisão e Revocação

| | |
|--------------|----------------|
| 1 - d_7 | 5 - d_{79} |
| 2 - d_6^* | 6 - d_{30} |
| 3 - d_2 | 7 - d_1^* |
| 4 - d_{13} | 8 - d_{15}^* |

Documentos com * também pertencem a R , isto é, são relevantes

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |



Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(0\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(0\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(10\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(0\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(10\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(20\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(0\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(10\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(20\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(30\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(0\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(10\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(20\%) = \max(p(25\%), p(50\%), p(75\%)) = \max(50\%, 29\%, 38\%) = 50\%$$

$$p(30\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

Apenas as precisões das revocações maiores ou iguais a de interesse é que devem ser consideradas (no caso, revocações maiores ou iguais a 30%)!

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(40\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(40\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(50\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(40\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(50\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(60\%) = \max(p(75\%)) = \max(38\%) = 38\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(40\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(50\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(60\%) = \max(p(75\%)) = \max(38\%) = 38\%$$

Apenas as precisões das revocações maiores ou iguais a de interesse é que devem ser consideradas (no caso, revocações maiores ou iguais a 60%)!

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(40\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(50\%) = \max(p(50\%), p(75\%)) = \max(29\%, 38\%) = 38\%$$

$$p(60\%) = \max(p(75\%)) = \max(38\%) = 38\%$$

Apenas as precisões das revocações maiores ou iguais a de interesse é que devem ser consideradas (no caso, revocações maiores ou iguais a 60%)!

$$p(70\%) = \max(p(75\%)) = \max(38\%) = 38\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(80\%) = 0\%$$

$$p(90\%) = 0\%$$

$$p(100\%) = 0\%$$

Precisão e Revocação

Temos assim a tabela inicial:

| Revocação | Precisão |
|-----------|----------|
| 25% | 50% |
| 50% | 29% |
| 75% | 38% |

Podemos então aplicar a regra de interpolação para calcular a precisão em cada um dos níveis padrão de revocação:

$$p(r_j) = \max_{r_k \geq r_j} (p(r_k))$$

$$p(80\%) = 0\%$$

$$p(90\%) = 0\%$$

$$p(100\%) = 0\%$$



Apenas as precisões das revocações maiores ou iguais a de interesse é que devem ser consideradas. Como não há, consideramos as precisões como zero.

Precisão e Revocação

Temos então a seguinte tabela com as precisões nos níveis de revocação padrão:

| Revocação | Precisão |
|-----------|----------|
| 0% | 50% |
| 10% | 50% |
| 20% | 50% |
| 30% | 38% |
| 40% | 38% |
| 50% | 38% |
| 60% | 38% |
| 70% | 38% |
| 80% | 0% |
| 90% | 0% |
| 100% | 0% |

Precisão e Revocação

Temos então a seguinte tabela com as precisões nos níveis de revocação padrão:

| Revocação | Precisão |
|-----------|----------|
| 0% | 50% |
| 10% | 50% |
| 20% | 50% |
| 30% | 38% |
| 40% | 38% |
| 50% | 38% |
| 60% | 38% |
| 70% | 38% |
| 80% | 0% |
| 90% | 0% |
| 100% | 0% |

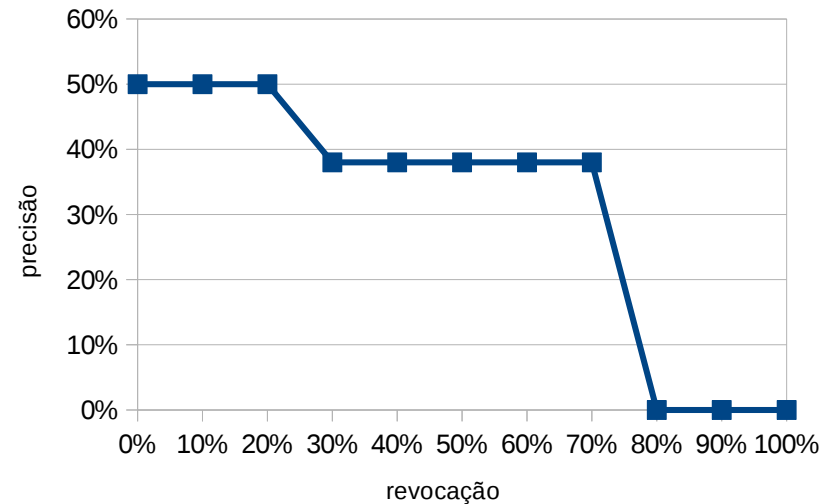
Por fim, geramos o gráfico de precisão e revocação:

Precisão e Revocação

Temos então a seguinte tabela com as precisões nos níveis de revocação padrão:

| Revocação | Precisão |
|-----------|----------|
| 0% | 50% |
| 10% | 50% |
| 20% | 50% |
| 30% | 38% |
| 40% | 38% |
| 50% | 38% |
| 60% | 38% |
| 70% | 38% |
| 80% | 0% |
| 90% | 0% |
| 100% | 0% |

Por fim, geramos o gráfico de precisão e revocação:



Precisão e Revocação

Comparando o gráfico inicial com o final:

Gráfico das revocações iniciais:

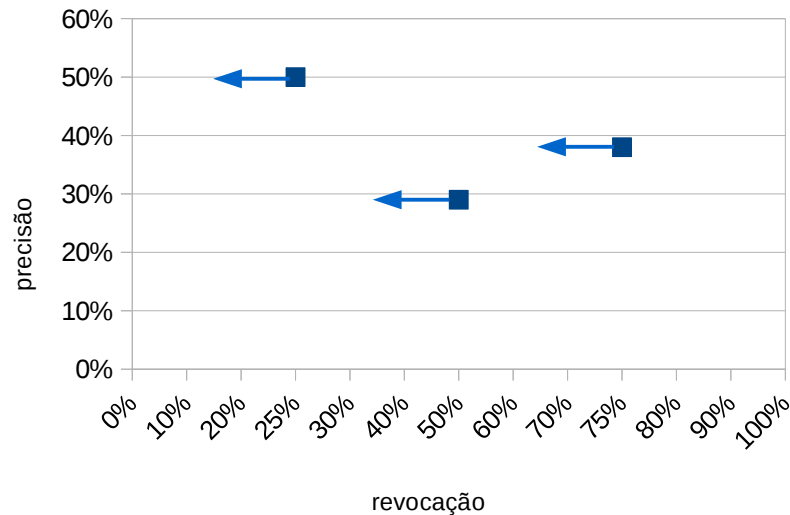
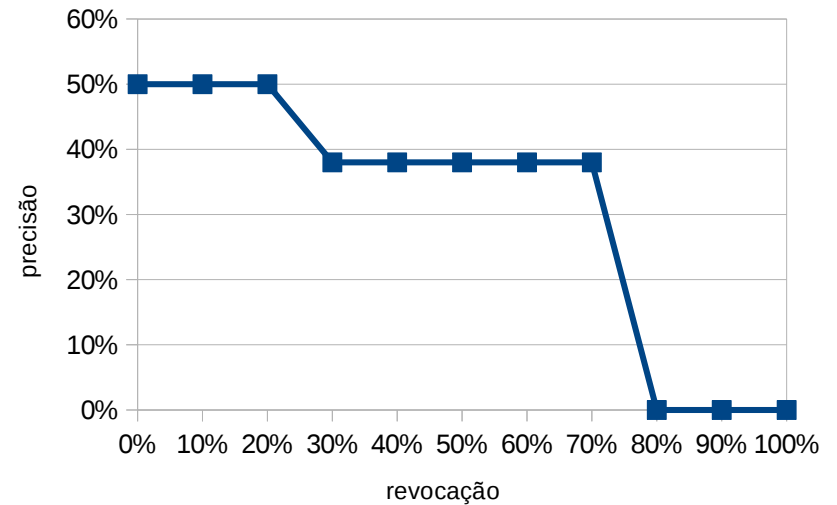


Gráfico final (nos níveis padrão de revocação):



Precisão e Revocação

- Como uma só consulta não é suficiente para avaliar o desempenho de um sistema, na prática, toma-se um conjunto de N_q consultas de teste e calcula-se a média das precisões $\bar{p}(r_j)$ para cada nível de revocação r_j :

$$\bar{p}(r_j) = \sum_{i=1}^{N_q} \frac{p_i(r_j)}{N_q}$$

Precisão na consulta i para nível de revocação r_j

Nº total de consultas de teste

Precisão e Revocação

- É comum comparar diferentes sistemas plotando suas curvas de revocação-precisão média no mesmo gráfico;
- Em alguns casos, costuma-se adotar a área abaixo da curva (AVC) como medida para se comparar os sistemas. Valores maiores para a área indicam maior qualidade;
- Empiricamente, tem-se observado que, ao se aumentar o nível de revocação, o nível de precisão diminui;
- Algoritmos com maiores níveis de precisão costumam ser preferíveis para a WEB. Em contextos mais específicos, como área médica ou jurídica, níveis maiores de revocação podem ser preferíveis.

Precisão e Revocação

Exercício: Considere uma coleção de referência. 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. Os conjuntos R1, R2 e R3 são dados da seguinte forma:

- $R1 = \{d3, d7, d12, d13, d26, d68\}$
- $R2 = \{d1, d2, d9, d24, d51, d52, d70, d82\}$
- $R3 = \{d2, d3, d6, d16, d20\}$

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 = {d1, d9, d26, d15, d2, d10, d74, d68, d32, d3, d53, d39, d56, d11, d4}.

Consulta q2 = {d3, d7, d8, d9, d19, d16, d37, d24, d20, d80, d67, d50, d46, d51, d29}.

Consulta q3 = {d2, d30, d25, d3, d9, d7d6, d39, d75, d19, d26 d16, d20, d51, d1}.

Precisão e Revocação

Exercício (continuação):

- a) Calcule os níveis de precisão e revocação para cada uma das consultas.
- 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.