



INSTITUTO FEDERAL
MINAS GERAIS

RECUPERAÇÃO DE INFORMAÇÃO

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



ATENÇÃO!!!

↴ O material a seguir é uma videoaula apresentada pela professora PATRÍCIA APARECIDA PROENÇA AVILA, como material pedagógico do IFMG, dentro de suas atividades curriculares ofertadas em ambiente virtual de aprendizagem. Seu uso, cópia e ou divulgação em parte ou no todo, por quaisquer meios existentes ou que vierem a ser desenvolvidos, somente poderá ser feito, mediante autorização expressa deste docente e do IFMG. Caso contrário, estarão sujeitos às penalidades legais vigentes”.

↴ Conforme Art. 2º§1º da Nota Técnica nº
1/2020/PROEN/Reitoria/IFMG (SEI 0605498, Processo nº
23208.002340/2020-04

Avaliação da Recuperação

Roteiro

- ▶ Introdução;
- ▶ Paradigma de Cranfield;
- ▶ Métricas de recuperação;
- ▶ Adequação das métricas.

Introdução

- ▶ Para avaliar um sistema de RI, é necessário medir o quão bem o sistema atende a necessidade de informação do usuário;
 - ▶ Isso pode ser um problema, pois um mesmo conjunto resposta pode ser interpretado de maneiras diferentes por usuários distintos;
- ▶ Mesmo assim, é possível definir métricas aproximadas, que, na média, têm uma correlação com as preferências de uma população.

Introdução

- ▶ Sem uma avaliação adequada, não temos como saber como o sistema de RI está desempenhando, nem podemos comparar objetivamente a qualidade e recuperação com a de outros sistemas.

Introdução

- ▶ Assim a avaliação sistemática de um sistema de RI permite responder a questões que surgem na prática durante sua manutenção:
 - i. Uma modificação na função de ranqueamento é proposta: devemos ir adiante e implantá-la?
 - ii. Uma nova função de ranqueamento probabilística foi projetada: ela é superior ao ranqueamento do modelo vetorial?
 - iii. Para quais tipos de consultas à Web, como consultas sobre negócios, produtos ou consultas geográficas, uma modificação no ranking funciona melhor?
- ▶ A falta de avaliação adequada impede que essas questões sejam respondidas de forma objetiva e impossibilita o ajuste da função de ranqueamento.

Introdução

- ▶ **Avaliação da recuperação** é um processo sistemático no qual se associa uma métrica quantitativa aos resultados produzidos por um sistema de RI em resposta a um conjunto de consultas de usuário.
- ▶ Essa métrica deve ser diretamente associada à relevância dos resultados para os usuários. Uma abordagem comum para computar tal métrica é **comparar o resultado produzido pelo sistema com os resultados sugeridos por humanos para o mesmo conjunto de consultas.**

Introdução

- ▶ Avaliação da recuperação significa avaliar a **qualidade dos resultados**, não o **desempenho** do sistema;
- ▶ A definição anterior não cobre aspectos que afetam o julgamento do usuário:
 - ▶ Interface do usuário;
 - ▶ Contexto da busca (preferências, período, etc) ;

Paradigma de Cranfield

- ▶ Em 1952, Cyril Cleverdon, bibliotecário da escola de aeronáutica de Cranfield, Inglaterra, conheceu o sistema Uniterm proposto por Mortimer Taube, bibliotecário nos EUA;
- ▶ Uniterm - 40 mil títulos formados por 7 mil palavras distintas;
 - ▶ O nome do sistema - os documentos eram indexados utilizando somente esses termos (palavras).
- ▶ Cleverdon achou o sistema interessante, indexou 200 documentos manualmente usando o sistema Uniterm e pediu para um colega que fizesse diversas consultas.

Paradigma de Cranfield

- ▶ Sistemas Uniterm x Sistemas de indexação mais sofisticados: briga entre os defensores de cada um deles;
 - ▶ pois o Uniterm parecia simples e não contemplava a semântica toda dos documentos.
 - ▶ mas não havia dados concretos que permitissem uma comparação direta entre os sistemas.
- ▶ Então Cleverdon propôs um estudo para comparar os diversos sistemas de indexação existentes;
 - ▶ Cada consulta era baseada em um só documento e a busca era considerada bem-sucedida se aquele documento fosse localizado no catálogo;

Paradigma de Cranfield

▶ Como foi realizado:

- ▶ Indexação manual de 18000 artigos sobre Engenharia Aeronáutica e avaliação de 1200 consultas;
- ▶ Utilização de quatro sistemas de indexação;
- ▶ Os resultados mostraram que os quatro sistemas de indexação eram basicamente equivalentes em termos da **precisão** dos resultados;

▶ Além disso os experimentos mostraram constatações interessantes:

- ▶ A **revocação** (fração dos documentos relevantes recuperados) era pouco útil se não estivesse acompanhada pela informação da **precisão** (fração dos documentos recuperados que é de fato relevante).
- ▶ Mas posteriormente experimentos mostraram que há um relacionamento inverso entre precisão e revocação, ou seja, não é possível aumentar simultaneamente a precisão e a revocação.

Paradigma de Cranfield

- ▶ Segundo experimento:
- ▶ 1400 documentos e 279 consultas;
- ▶ Para cada consulta, todos os documentos da coleção foram examinados e sua relevância em relação à consulta foi determinada;
- ▶ Seis estudantes passaram três meses examinando cada documento em relação a consulta e decidindo quais documentos eram relevantes;
- ▶ O resultado foi uma coleção de referência.
 - ▶ Composta por: documentos, consultas e julgamentos de relevância de cada par consulta-documento.
 - ▶ Coleção ficou conhecida como Cranfield-2.

Paradigma de Cranfield

- ▶ Nos experimentos foi observado também:
 - ▶ Em situações práticas, a maioria das buscas não requer uma revocação (recuperação) alta;
 - ▶ A maioria dos usuários requer apenas algumas respostas relevantes (precisão);
- ▶ Os experimentos Cranfield 2 estabeleceram a base para a **experimentação moderna em RI**:
 - ▶ O mesmo conjunto de documentos e consultas pode ser usado para avaliar sistemas de ranqueamento diferentes comparando-os com os julgamentos de relevância produzidos por especialistas humanos.
 - ▶ Além disso, **precisão e revocação** são hoje as métricas escolhidas para avaliar a qualidade do ranking.

Coleção de referência

- ▶ Coleções de referência permitem comparar diretamente os resultados produzidos por diferentes funções de ranqueamento;
- ▶ Coleção de referência: é composta por um conjunto D de documentos pré-selecionados, um conjunto I de necessidade de informação usadas para teste e um conjunto de julgamentos de relevância binários associados a cada par $[i_m, d_j]$.

Coleção de referência

- ▶ Os julgamentos de relevância são produzidos por humanos especialistas e idealmente devem fornecer uma decisão de relevância para cada par necessidade de informação-documento.
- ▶ Claramente, isso só é viável para coleções de documento pequenas, como as dos experimentos Cranfield;
- ▶ Os julgamentos de relevância são fornecidos para as descrições das necessidades de informação, e não para as consultas.
 - ▶ traduzir a necessidade de consulta faz parte do sistema de RI, e cada um trata de uma maneira.

Coleção de referência

► Vantagens:

- Dada uma coleção de referência, uma avaliação do sistema de RI pode ser feita rapidamente;
- Avaliações de sistemas de RI podem ser reproduzidas posteriormente para fins de verificação (Repetitividade);
- Coleções de referência diferentes podem ser construídas focando em tipos particulares de necessidade de informação.

Métricas de Recuperação

- ▶ Na literatura podemos encontrar várias métricas para a avaliação da qualidade da recuperação de um sistema de RI, isto é a qualidade dos resultados.
- ▶ As mais amplamente utilizadas são **precisão** e **revocação**.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**

- ▶ As medidas de precisão e revocação são definidas da seguinte forma:

- ▶ **Precisão** (fração dos documentos recuperados que é relevante):

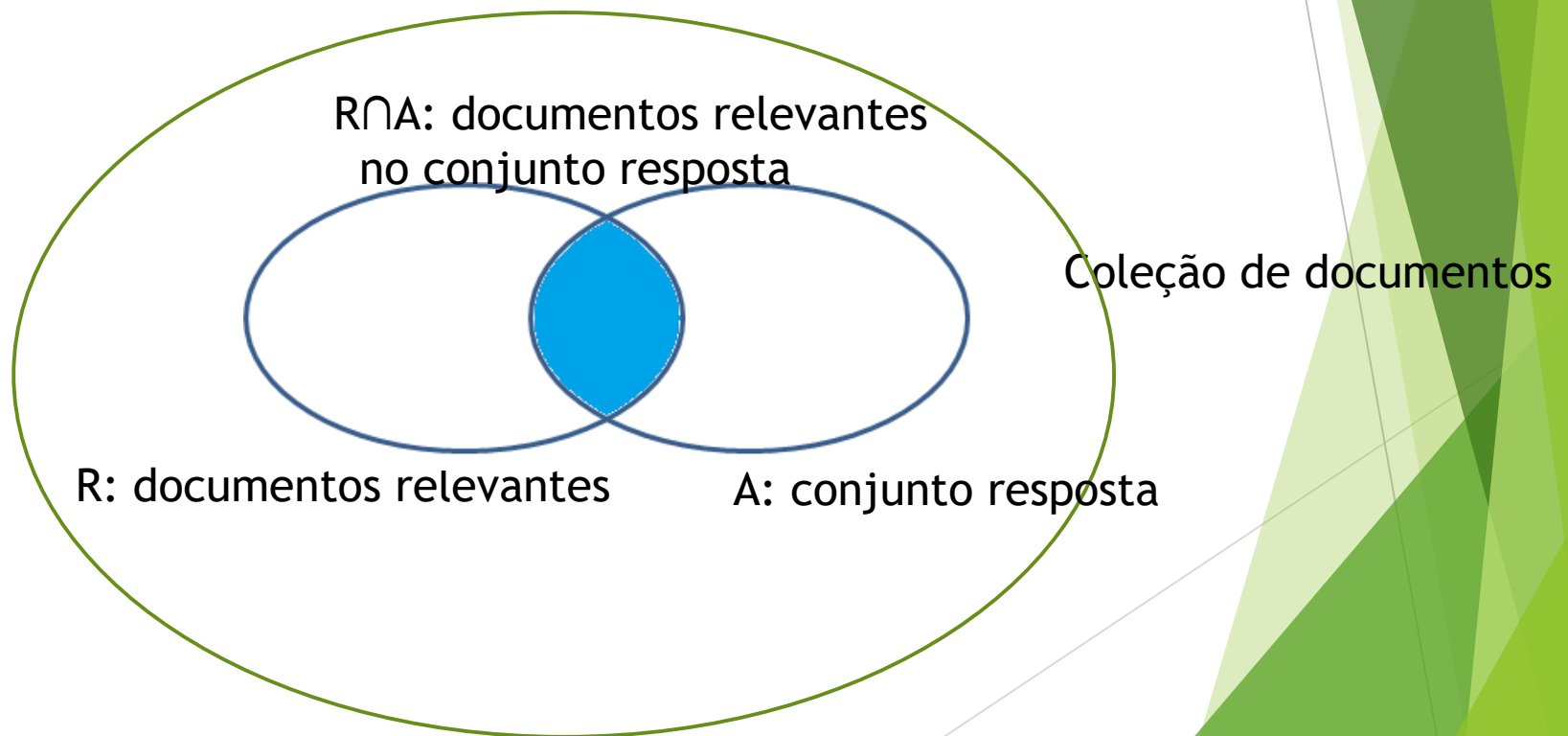
$$p = |R \cap A| / |R|$$

- ▶ **Revocação** (fração dos documentos relevantes que foi recuperada):

$$r = |R \cap A| / |A|$$

Métricas de Recuperação

► Precisão e Revocação:



Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ Tais métricas supõem que todos os documentos no conjunto resposta A foram examinados;
- ▶ Contudo, o usuário não é normalmente apresentado a todos os documentos do conjunto resposta A de uma vez;
- ▶ Em vez disso, os documentos do conjunto A são primeiro ordenados de acordo com uma função de ranqueamento;
- ▶ O usuário examina essa lista ordenada iniciando pelo documento do topo:
 - ▶ Nesse caso, as medidas de precisão e revocação variam conforme o usuário procede com seu exame do conjunto resposta.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ Portanto, uma avaliação adequada requer a plotagem da curva de precisão x revocação.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**

- ▶ **Exemplo 1:**

- ▶ Coleção referência e um conjunto de consultas para teste.
- ▶ Supondo que R_1 - conjunto de documentos relevantes para uma dada consulta q_1 (determinado por especialistas).
 - ▶ $R_1 = \{d3, d5, d9, d25, d39, d44, d56, d71, d89, d123\}$

Métricas de Recuperação

- **Precisão e Revocação:**

- **Exemplo 1:**

- Supondo um novo algoritmo de recuperação que retorne para a consulta q_1 o ranking:

1. d123	6. d9	11. d38
2. d84	7. d511	12. d48
3. d56	8. d129	13. d250
4. d6	9. d187	14. d113
5. d8	10. d25	15. d3

- Os documentos relevantes para a consulta estão em negrito.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**

- ▶ **Exemplo 1:**

- ▶ Primeiro d123: um documento relevante ($R \cap A$) e um documento recuperado (A)

- ▶ $p = 1/1 = 100\%$

- ▶ $r = 1/10 = 10\%$

- ▶ Segundo d56: dois documentos relevantes ($R \cap A$) e três documentos recuperados (A)

- ▶ $p = 2/3 = 66,67\%$

- ▶ $r = 2/10 = 20\%$

- ▶ Terceiro d9: três documentos relevantes ($R \cap A$) e seis documentos recuperados (A)

- ▶ $p = 3/6 = 50\%$

- ▶ $r = 3/10 = 30\%$

Métricas de Recuperação

- ▶ **Precisão e Revocação:**

- ▶ **Exemplo 1:**

- ▶ Quarto d25: quatro documentos relevantes ($R \cap A$) e dez documentos recuperados (A)

- ▶ $p = 4/10 = 40\%$

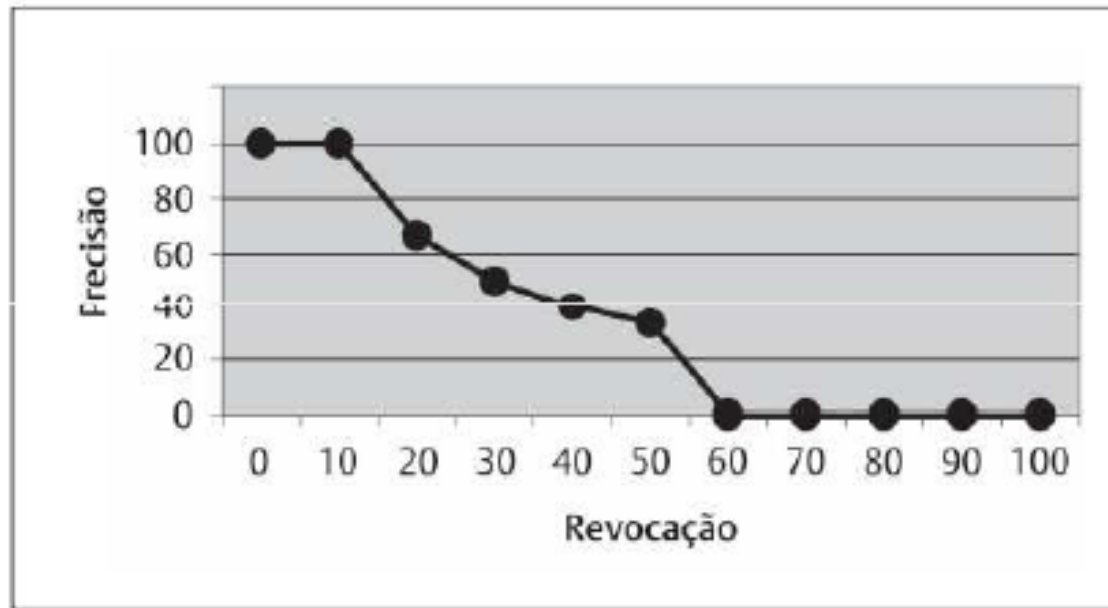
- ▶ $r = 4/10 = 40\%$

- ▶ Quinto d3: cinco documentos relevantes ($R \cap A$) e quinze documentos recuperados (A)

- ▶ $p = 5/15 = 33,3\%$

- ▶ $r = 5/10 = 50\%$

Métricas de Recuperação



Revocação	Precisão
0	100
10	100
20	66,6
30	50
40	40
50	33,3
60	0
70	0
80	0
90	0
100	0

* A precisão para os níveis de revocação maiores que 50% cai para 0 porque nem todos os documentos relevantes foram recuperados.

* Revocação 0 é obtida por meio de interpolação.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ **Exemplo 2:**
 - ▶ Considere uma segunda consulta q_2 , cujo conjunto de documentos relevantes, de acordo com especialistas humanos, seja:
 - ▶ $R_2 = \{d3, d56, d129\}$

Métricas de Recuperação

- **Precisão e Revocação:**

- **Exemplo 2:**

- Supondo o mesmo algoritmo de recuperação do exemplo anterior, retorne para a consulta q_2 o ranking:

1. d425	6. d615	11.d193
2. d87	7. d512	12.d715
3. d56	8. d129	13.d810
4. d32	9. d4	14.d5
5. d124	10.d130	15.d3

- Os documentos relevantes para a consulta estão em negrito.

Métricas de Recuperação

- ▶ **Precisão e Revocação:**

- ▶ **Exemplo 2:**

- ▶ Primeiro d56: um documento relevante ($R \cap A$) e 3 documentos recuperado (A)
 - ▶ $p = 1/3 = 33,3\%$
 - ▶ $r = 1/3 = 33,3\%$
- ▶ Segundo d129: dois documentos relevantes ($R \cap A$) e oito documentos recuperados (A)
 - ▶ $p = 2/8 = 25\%$
 - ▶ $r = 2/3 = 66,6\%$
- ▶ Terceiro d3: três documentos relevantes ($R \cap A$) e quinze documentos recuperados (A)
 - ▶ $p = 3/15 = 20\%$
 - ▶ $r = 3/3 = 100\%$

Métricas de Recuperação

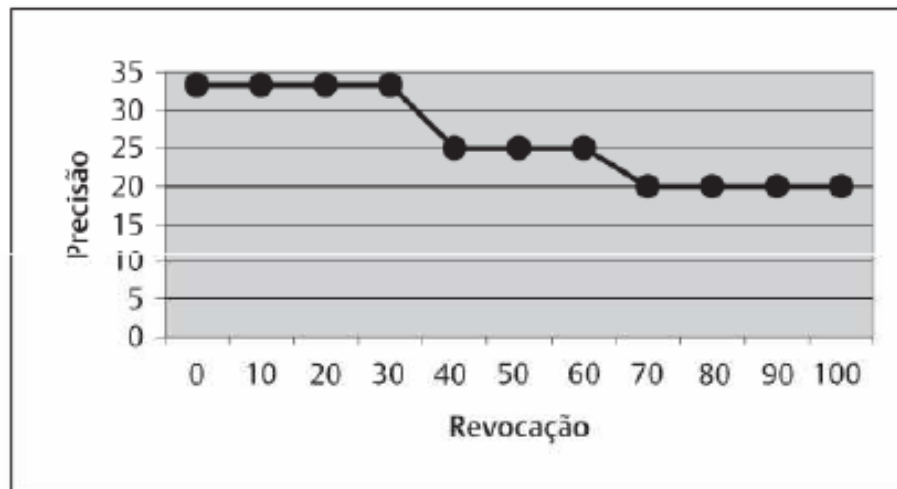
- ▶ **Precisão e Revocação:**
- ▶ Precisão nos 11 níveis padrão de revocação:
 - ▶ Utiliza-se um método de interpolação para se obter a precisão em 11 pontos de revocação (0%, 10%, 20%, ..., 100%)
 - ▶ Seja r_j , $j \in \{0, 1, 2, \dots, 10\}$, uma referência ao j -ésimo nível padrão de revocação (isto é, r_5 é uma referência ao nível de revocação 50%).
 - ▶ $P(r_j) = \max P(r)$
 - ▶ Ou seja, a precisão máxima conhecida entre todos os níveis de revocação maior ou igual a r_j .

Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ **Exemplo 2:**
 - ▶ Voltando ao exemplo para calcular os 11 pontos de revocação.
 - ▶ Nos níveis de revocação 0%, 10%, 20% e 30%, a precisão interpolada é 33,3%, pois a revocação conhecida é 33,3% que é maior ou igual do que esses pontos.
 - ▶ Nos níveis de revocação 40%, 50% e 60%, a precisão interpolada é 25%, pois a revocação conhecida é 66,6% que é maior ou igual a esses pontos.
 - ▶ Nos demais níveis, a precisão é 20%, conforme analisado anteriormente.

Métricas de Recuperação

- Precisão e Revocação:
- Exemplo 2:



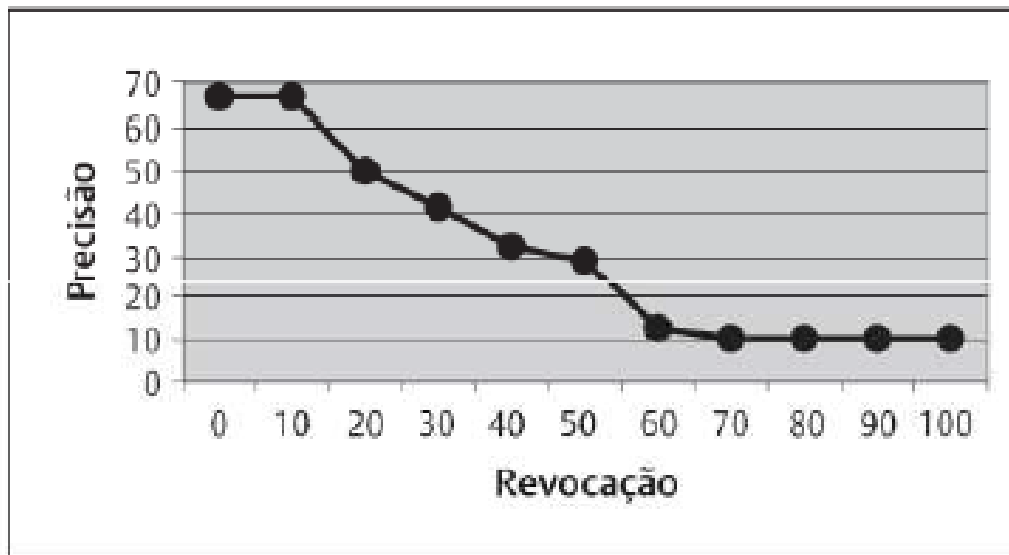
Revocação	Precisão
0	33,3
10	33,3
20	33,3
30	33,3
40	25
50	25
60	25
70	20
80	20
90	20
100	20

Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ Geralmente os algoritmos são avaliados sobre diversas consultas de teste. Neste caso uma curva de precisão x revocação é gerada para cada consulta.
- ▶ Para avaliar a qualidade de recuperação de um algoritmo sobre um conjunto de N consultas, calculamos a média das precisões para cada nível de revocação.

Métricas de Recuperação

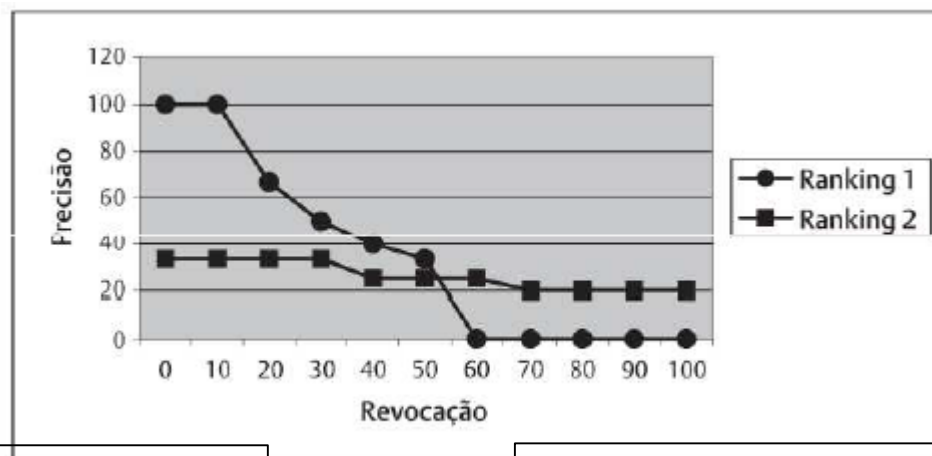
- Precisão e Revocação:
- Exemplo (média entre q1 e q2)



Revocação	Precisão
0	66,6
10	66,6
20	49,9
30	41,6
40	32,5
50	29,1
60	12,5
70	10
80	10
90	10
100	10

Métricas de Recuperação

- ▶ **Precisão e Revocação:**
- ▶ As médias dos valores de precisão e revocação podem também se utilizadas para comparar diferentes algoritmos de recuperação.



Primeiro: voltados para a busca Web, pois devem apresentar níveis de precisão mais altos.

Segundo: aplicações específicas como as da área da saúde ou jurídica necessitam de níveis altos de revocação.

Adequação das Métricas

- ▶ Precisão e revocação têm sido amplamente utilizadas para avaliar a qualidade de algoritmos de recuperação;
 - ▶ Uma reflexão mais cuidadosa revela problemas com essas medidas;
- ▶ Veremos cada um deles.

Adequação das Métricas

► Problemas:

1. A estimativa da **revocação máxima** para uma consulta requer um conhecimento detalhado de todos os documentos da coleção - para coleções grandes tal conhecimento não está disponível;
2. Precisão e revocação são medidas relacionadas que capturam diferentes aspectos do conjunto de documentos recuperados - em certos casos uma só medida pode ser mais apropriada;
3. Precisão e revocação medem a eficácia para um conjunto de consultas processadas em lotes - sistemas modernos a interatividade é um aspecto fundamental do processo de recuperação.

Adequação das Métricas

- ▶ Apesar dessas desvantagens, precisão e revocação continuam sendo amplamente utilizadas, porque, dada uma coleção de referência, elas são simples de calcular e permitem a comparação direta de diferentes estratégias de ranqueamento.

Conclusão

- ▶ Como avaliar um modelo de recuperação da informação usando uma coleção de referência;
- ▶ Duas métricas amplamente utilizadas foram apresentadas: precisão e revocação.
 - ▶ **Precisão** - representa a fração de documentos recuperados que são relevantes;
 - ▶ **Revocação** - representa a fração de documentos relevantes que foram recuperados.

ATIVIDADES PARA ENTREGAR - 02/08

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

ATIVIDADES PARA ENTREGAR - 02/08

- ▶ 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}.*

ATIVIDADES PARA ENTREGAR - 02/08

- ▶ 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.

Onde não há amor,
coloca amor, e colherá
amor. (São João da Cruz)

